



**U.S. Army Research Institute
for the Behavioral and Social Sciences**

Research Report 2013

**Decision Environment and Heuristics in Individual and
Collective Hypothesis Generation**

**Drew A. Leins
Jim Leonard
Laura A. Zimmerman**
Applied Research Associates

Maria L. Minchew
Engility Corporation

Christopher L. Vowels
U.S. Army Research Institute

November 2017

Approved for public release; distribution is unlimited.

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

**Department of the Army
Deputy Chief of Staff, G1**

Authorized and approved for distribution:

**MICHELLE SAMS
Director**

Technical review by

Dr. Victor Ingurgio, U.S. Army Research Institute
Dr. LisaRe Babin, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Research Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-ZXM, 6000 6th Street, Bldg 1464/Mail Stop 5610, Fort Belvoir, VA 22060-5610.

FINAL DISPOSITION: This Research Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)			2. REPORT TYPE Final			3. DATES COVERED (from... to) March 2015 – March 2016		
4. TITLE AND SUBTITLE Decision Environment and Heuristics in Individual and Collective Hypothesis Generation						5a. CONTRACT OR GRANT NUMBER W5J9CQ-11-0003-0005		
						5b. PROGRAM ELEMENT NUMBER 622785		
6. AUTHOR(S) Drew A. Leins, Jim Leonard, and Laura Zimmerman (Applied Research Associates), Maria L. Minchew (Engility Corporation), and Christopher L. Vowels (U.S. Army Research Institute)						5c. PROJECT NUMBER A790		
						5d. TASK NUMBER 215		
						5e. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Applied Research Associates Engility Corporation 2760 Eisenhower Avenue, Suite 308 4803 Stonecroft Boulevard Alexandria, VA 22314 Chantilly, VA 20151 U.S. Army Research Institute for the Behavioral and Social Sciences Bldg 10007 Old Ironsides Avenue Fort Hood, TX 76544						8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6 th Street, Bldg 1464/Mail Stop 5610 Fort Belvoir, VA 22060-5610						10. MONITOR ACRONYM ARI		
						11. MONITOR REPORT NUMBER Research Report 2013		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.								
13. SUPPLEMENTARY NOTES Contracting Officer Representative and Subject Matter POC: Dr. Christopher L. Vowels								
14. ABSTRACT (<i>Maximum 200 words</i>): In two experiments, time pressure, familiarity of the decision environment, onset of informative cues, and the level of certainty portrayed within the decision environment were manipulated while examining possible changes in hypotheses generated and confidence in those hypotheses in individual and collective conditions. Hypothesis generation was examined as a function of heuristic usage to better understand the possible effect of the manipulated factors on the production of situational assessments. In conditions with no time pressure, when the decision context was unfamiliar, and when Soldiers received only low-value information early in the scenario, Soldiers likely invoked a weighted-additive strategy or allowed maximal information to accumulate in scenarios before making a situational assessment. Soldiers working collectively demonstrated a faster generation of hypotheses when the decision environment was more familiar to them. Additionally, Soldiers working together were less likely to base their hypotheses on suboptimal information.								
15. SUBJECT TERMS Hypothesis, Hypothesis Generation, Training, Heuristic, Collective Performance, Squad								
SECURITY CLASSIFICATION OF						19. LIMITATION OF ABSTRACT Unlimited	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON Brian T. Crabb Unit Chief 254-288-3832
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified						

Research Report 2013

Decision Environment and Heuristics in Individual and Collective Hypothesis Generation

**Drew A. Leins
Jim Leonard
Laura A. Zimmerman**
Applied Research Associates

Maria L. Minchew
Engility Corporation

Christopher L. Vowels
U.S. Army Research Institute

**Fort Hood Research Unit
Brian T. Crabb, Chief**

U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street, Bldg 1464, Fort Belvoir, VA 22060-5610

November 2017

**Army Project Number
622785A790**

**Personnel, Performance
and Training Technology**

Approved for public release; distribution is unlimited.

ACKNOWLEDGEMENT

The authors would like to thank the Soldiers and leaders who took time to participate in this research and provide their insights and experiences involving individual and collective decision-making. We would also like to thank the unit leadership for finding time to support the data collections among several competing priorities.

DECISION ENVIRONMENT AND HEURISTICS IN INDIVIDUAL AND COLLECTIVE HYPOTHESIS GENERATION

EXECUTIVE SUMMARY

Research Requirement

Threat detection requires Soldiers to engage in several cognitive processes. They perceive and process situational cues, and generate and evaluate hypotheses, before determining whether a situation poses a threat. How individual Soldiers engage in these processes may impact how their squad performs collectively. Through exploring and measuring these processes, the Army can develop better methods for assessing and training associated skills. One goal of this research was to establish valid measures for some of the cognitive processes inherent in individual and collective hypothesis generation. A second goal of this research was to explore whether, and when, Soldiers employ cognitive mechanisms to generate hypotheses more efficiently. This report presents two experiments that explored influences of hypothesis generation among Soldiers performing individually (Experiment 1) and collectively in groups (Experiment 2).

Threat detection in an operational environment is ideal for using heuristics to guide decision-making (e.g., see Rieskamp and Hoffrage, 1999). When the cost of slow, effortful deliberation appears prohibitive (e.g., life threatening), it will be sacrificed for quicker, more efficient decision processes. Image theory (Beach, 1990) is useful for considering how threat cues in the environment may activate heuristics. Environmental cues activate schemas in memory (Thomas, Dougherty, Sprenger, & Harbison, 2008). As emerging cues alter perceptions of the environment, Soldiers may re-evaluate the perceived threat level of an environment. Thus, decision makers will generate hypotheses at the rate at which they recognize environmental cues and any subsequent changes in those cues. Strong correspondence with a particular schema in memory may lead to a quick hypothesis, suggesting that both familiarity and the order in which decision makers perceive environmental cues influence the speed with which they generate hypotheses. In addition, the time available for decisions may also influence whether decision makers use heuristics when generating hypotheses. Consider that heuristics can operate as adaptive stop rules. Critically, when decision makers do not recognize any cue as informative, they may continue searching the environment. This becomes problematic when search time is finite. In this case, decision makers can delay a decision or adjust the threshold for evaluating cues. Thomas, Dougherty, Sprenger, and Harbison (2008) suggest that decision makers adjust their criteria for generating hypotheses under time constraints. But, adjusting the threshold of informativeness can result in suboptimal hypotheses. The current research explored Soldiers' heuristic usage when generating threat hypotheses in familiar versus unfamiliar tasks, under time pressure or no time pressure, when informative cues appeared early versus late in a scenario, and when working individually or collectively.

Procedure

Experiments 1 & 2 used a 2 x 2 x 2 (cue-order [high-value first vs low-value first] x familiarity [familiar vs. unfamiliar] x time pressure [low vs. high]) fully-crossed, within-subjects design. Participants worked either individually (Experiment 1) or collectively (Experiment 2) to enter responses into a laptop computer. Thirty-three Soldiers were tested individually (Experiment 1) and 44 Soldiers working in groups of 3-4 (Experiment 2).

Participants generated hypotheses for 12 scenarios, half of which presented threat detection (familiar) tasks and half presented medical diagnosis (unfamiliar) tasks. Each scenario included a short description accompanied by an image. The descriptions presented the scenario context and specific decision requirements. A request for an assessment of the scenario then followed. After participants entered their assessment and confidence rating, a new cue was added to the image every six seconds until three new cues had been added or until the participant stopped the trial to indicate a change in assessment. Each scenario contained one high-value cue presented in either the second or third serial order position. After completing all scenarios, participants completed a demographic questionnaire and two decision-making disposition scales, the Decision-Making Style (Scott & Bruce, 1995) and the Need for Cognitive Closure (Roets & Van Hiel, 2011; for the original scale, see Webster & Kruglanski, 1994)

The initial and secondary hypotheses were scored for threat level and calculated the average response latency, quality of hypothesis timing, and confidence scores across scenarios within each condition. To account for group member contribution, the proportions of contribution were calculated for each participant in each scenario and used these proportion scores to calculate the contribution variance for each group across scenarios and conditions.

Experiment 1 Results

Soldiers reported low to moderate initial threat levels across scenarios. Soldiers tended to report increases in threat levels over time. Soldiers also reported greater increases in threat levels when experiencing no time pressure than when under time pressure.

Time pressure also influenced the number of images viewed. Soldiers reported changes in their hypotheses sooner (i.e., viewed fewer images) under time pressure versus no time pressure. Time pressure interacted with cue order interacted to influence number of images viewed: When the high-value cue appeared late in the trial, time pressure yielded hypothesis changes sooner than did no time pressure. However, when the high-value cue appeared early in the trial, time pressure had no effect on images viewed.

Overall, participants' timing was optimal (i.e., it corresponded with the high-value cue) in 46% of scenarios and suboptimal (i.e., it occurred prior to the high-value cue) in 54% of scenarios. Cue order influenced the quality of timing. Participants were more likely to change their hypotheses at an optimal time when the high-value cue appeared late versus when it appeared early.

Cue order nested within familiarity influenced the quality of timing. When completing unfamiliar scenarios, participants were more likely to change their hypotheses at a more optimal

time when the high-value cue appeared late versus early. When completing familiar scenarios, cue order did not influence the quality of timing. Time pressure also appeared to influence the quality of timing. Participants were more likely to change their hypotheses at an optimal time when under no perceived time pressure versus when under perceived time pressure.

Soldiers' Need for Cognitive Closure (NFCC) scores correlated positively with images viewed when under time pressure and with early presentation of the high-value cue. In both conditions, higher NFCC scores correlated with a greater number of images viewed.

Experiment 2 Results

Similar to individuals in Experiment 1, groups of Soldiers in Experiment 2 reported low to moderate initial threat levels and tended to reported increasing threat levels over time.

Familiarity influenced changes in reported threat level. Groups reported greater increases in threat levels for unfamiliar scenarios than for familiar scenarios. Time pressure and familiarity interacted to influence groups' reported changes in threat level: When completing familiar scenarios, groups reported larger increases in threat level when under no time pressure versus when under time pressure, whereas when completing unfamiliar scenarios, groups reported larger increases in threat level when under time pressure versus when under no time pressure.

Cue order influenced the number of images viewed. Groups viewed fewer images when the high-value cue appeared early versus when the high-value cue appeared late in the trial. Cue order also interacted with familiarity. In familiar contexts, groups viewed fewer images when high-value cues appeared early versus when they appeared late. By contrast, cue order had no effect on images viewed in unfamiliar contexts.

Group hypothesis timing was optimal in 57% of scenarios and suboptimal in 43% of scenarios in Experiment 2. Cue order influenced the quality of timing. Groups were more likely to change their hypotheses at an optimal time when the high-value cue appeared late versus when it appeared early. Cue order nested within familiarity also influenced the quality of timing. When completing familiar scenarios, groups were more likely to change their hypotheses at a more optimal time when the high-value cue appeared late versus early. Similarly, when completing unfamiliar scenarios, groups changed their hypotheses at more optimal times when the high-value cue appeared late versus early.

In Experiment 2, rank correlated positively with individual contribution. Higher ranking participants contributed more to group discussions. The amount of time a group spent working together in their respective units correlated negatively with the distribution of participant contribution within the group. As the amount of time spent together increased, contribution was more evenly distributed across group members.

In all conditions, groups of participants (Experiment 2) viewed more images than did individual participants (Experiment 1), all $t > 1.94$, all $p \leq .06$, all $d > 0.66$.

Utilization and Dissemination of Findings

Across experiments, influences of cue order, decision task familiarity, and time pressure on the number of images viewed before changing hypotheses were observed, suggesting that Soldiers engaged different hypothesis generation strategies as a function of the context or decision space in which they operated. In some contexts, Soldiers viewed all the images and thus, evaluated all possible information before reporting a new hypothesis. Across experiments, Soldiers may have adopted weighted-additive strategies when they experienced no time pressure, when the decision context was unfamiliar, and when they received only low-value information early in the scenario. These conditions presented ambiguous information early – thus offering no obvious disparity in cue values – and they presented no consequence for waiting for more valuable information. In efforts to gain certainty in these scenarios, Soldiers simply waited for more information.

Alternatively, Soldiers may have employed heuristics across all experimental conditions, but this manifested only when Soldiers encountered conditions that altered how they applied heuristics. Under increased time pressure, a familiar context, and early access to valuable information Soldiers may have modified satisficing heuristics already in use, to make them more efficient. Specifically, time pressure combined with a delay in receiving critical information seemed to induce individual Soldiers to lower their criterion for judging the informativeness of environmental cues and to trigger new hypotheses. It was also observed that familiar decision environments promoted quicker hypothesis generation among groups of Soldiers. These groups seemingly leveraged recognition-based heuristics in a way that individual Soldiers did not.

These effects suggest that under certain conditions, Soldiers may generate hypotheses informed by suboptimal information. Encouraging, however, is that multiple factors can mitigate potentially faulty hypothesis generation. First, individual disposition can influence whether Soldiers working individually employ heuristics. Soldiers high in the need for cognitive closure may be less susceptible to engaging in quick, efficient, but risky hypothesis generation. In addition, working in groups may mitigate the tendency to employ heuristics. Soldiers working in groups appeared less likely than individuals to base their hypotheses on suboptimal information. To be sure, they were generally less efficient than individual Soldiers, but the influence of group dynamic and group decision processes may have protected Soldiers against satisficing and potentially suboptimal hypothesis generation.

The experiments presented here represent one step toward understanding how decision environments influence the way Soldiers use heuristics to generate hypotheses individually and collectively. The findings imply that as Soldiers perceive their environments and the demands of their tasks differently, they may also assess those environments differently. Different assessments can lead to considering different courses of action that can directly impact Soldier safety and mission success. Therefore, it is critical to better understand the relative influences of hypothesis generation in operational environments. In addition to exploring the factors addressed here, future research should explore the influence of training and experience on the relationships between environmental conditions, decision tasks, and cognitive processes.

DECISION ENVIRONMENT AND HEURISTICS IN INDIVIDUAL AND COLLECTIVE HYPOTHESIS GENERATION

CONTENTS

	Page
Hypothesis Generation in Threat Detection.....	1
Heuristics in Hypothesis Generation	2
Experiment 1: Exploring Hypothesis Generation at the Individual Level.....	5
Design	5
Independent Measures	5
Dependent Measures	6
Experimental Hypotheses	6
Method	7
Participants.....	7
Materials	7
Procedure	9
Scoring	9
Results.....	10
Initial hypothesis threat-level.....	10
Changes in hypothesis threat-level	10
Images viewed	11
Quality of timing.....	13
Initial Confidence.....	14
Change in confidence.....	15
Decision-Making Scales	16
Discussion	17
Experiment 2: Exploring Hypothesis Generation at the Collective Level.....	20
Design	22
Independent Measures	22
Dependent Measures	22
Experimental hypotheses	22
Method	22
Participants.....	22
Materials	23
Procedure	23
Scoring	24
Results.....	24
Initial hypothesis threat-level.....	24
Changes in hypothesis threat-level	24
Images viewed	25
Quality of timing.....	27

CONTENTS (Continued)

	Page
Initial confidence	28
Change in confidence.....	28
Group specific factors	30
Individual versus group participants	30
Discussion	31
General Discussion	34
References	37
Appendix A. Scenario Examples	A-1
Appendix B. Questionnaires and Scales	B-1
Appendix C. Hypothesis Scoring Examples	C-1

LIST OF TABLES

Table 1. Experiment 1 Sample Demographics	7
Table 2. Threat Level Change: Descriptive Statistics for Cue Order Nested within Familiarity	11
Table 3. Images Viewed: Descriptive Statistics for Cue Order Nested within Familiarity	12
Table 4. Images Viewed: Descriptive Statistics for Time Pressure x Familiarity	12
Table 5. Images Viewed: Descriptive Statistics for Time Pressure x Cue Order	13
Table 6. Quality of Timing: Descriptive Statistics for Cue Order Nested in Familiarity	14
Table 7. Quality of Timing: Descriptive Statistics for Time Pressure x Cue Order	14
Table 8. Confidence: Descriptive Statistics for Familiarity x Time Pressure.....	15
Table 9. Confidence Change: Descriptive Statistics for Cue Order Nested in Familiarity	16
Table 10. Confidence Change: Descriptive Statistics for Time Pressure x Familiarity	16
Table 11. Confidence Change: Descriptive Statistics for Time Pressure x Cue Order	16
Table 12. Need for Cognitive Closure Scale Scores Association with Images Viewed.....	17
Table 13. Experiment 2 Sample Demographics	23
Table 14. Threat Level Change: Descriptive Statistics for Time Pressure x Familiarity	25
Table 15. Images Viewed: Descriptive Statistics for Cue Order Nested in Familiarity	26
Table 16. Images Viewed: Descriptive Statistics for Time Pressure x Familiarity	26
Table 17. Images Viewed: Descriptive Statistics for Time Pressure x Cue Order	26
Table 18. Quality of Timing: Descriptive Statistics for Cue Order Nested in Familiarity	27
Table 19. Confidence: Descriptive Statistics for Familiarity x Time Pressure.....	28
Table 20. Confidence Change: Descriptive Statistics for Cue Order Nested in Familiarity	29
Table 21. Confidence Change: Descriptive Statistics for Time Pressure x Familiarity	29
Table 22. Confidence Change: Descriptive Statistics for Time Pressure x Cue Order	29
Table 23. Images Viewed: Effect of Individual versus Group	31
Table 24. Quality of Timing: Effect of Individual versus Group	31

DECISION ENVIRONMENT AND HEURISTICS IN INDIVIDUAL AND COLLECTIVE HYPOTHESIS GENERATION

The ability to train and measure collective performance is critical to ensuring Army mission success. Underlying collective performance are individual cognitive processes. For example, in the context of threat detection, which comprises both individual and collective components, Soldiers engage in several cognitive processes. They perceive and process situational cues, and generate and evaluate hypotheses, before determining whether a situation poses a threat. How individual Soldiers engage in these processes may impact how their squad performs collectively. For example, different Soldiers may process situational cues with differing efficiency. To date, no empirical research has addressed the influence of individual cognitive processes on collective performance in threat detection. Such research will help develop valid measures of cognitive processes, which the Army can transform into better methods for assessing and training associated skills. Thus, one goal of the research reported here was to support the development of a valid framework for measuring cognitive processes inherent in individual and collective hypothesis generation. Cognitive mechanisms that increase efficiency – for example, processes that reduce the amount of information required to generate a valid hypothesis – may also enhance performance, particularly on cognitively complex tasks such as threat detection. Thus, a second goal of this research was to explore whether, and when, Soldiers employ such cognitive mechanisms while generating hypotheses. This report presents two experiments that explored influences of hypothesis generation among Soldiers performing individually (Experiment 1) and collectively (Experiment 2).

Hypothesis Generation in Threat Detection

When decision makers generate hypotheses, they process a set of environmental cues to create candidate explanations of the environment (e.g., Fisher, Gettys, Manning, Mehle, & Baca, 1983). Across a variety of decision contexts, decision makers tend to generate hypotheses with little conscious effort (Thomas, Dougherty, Sprenger, & Harbison, 2008). That is, cue perception and hypothesis generation often occur at a level below conscious awareness. One type of model accounting for this type of hypothesis generation is the recognition-based model (e.g., the Recognition-Primed Decision model; Klein, 1993, 1997). In recognition-based models, decision makers identify critical cues in the environment and match those cues to details in memory of previously experienced situations (recognition). Such recognition serves to recall previously evaluated, tested, and validated hypotheses. This allows decision makers to choose a course of action and predict an outcome swiftly, certainly, and confidently. However, successfully employing this type of hypothesis generation generally requires two conditions: familiar environments and experienced decision makers. By contrast, when environments are unfamiliar or decision makers are inexperienced, recognition may not occur and decision makers must either generate new hypotheses or offer existing hypotheses (or entertain a combination of both new and existing hypotheses) for evaluation and testing. These alternatives may require effortful evaluation, thus slowing decision-making; and, they may suffer from weaker associations in memory, thus affording less certainty and less confidence to decision makers. Research tends to demonstrate quantitative differences in decision-making as a function of environment familiarity and decision maker experience. As environments become more familiar and decision makers accrue more patterns in memory, they recognize critical cues more quickly

and rely on fewer cues to generate hypotheses (Cohen, Freeman, & Wolf, 1996; Klein, 1993; Klein, 1997; Lipshitz & Strauss, 1997).

Detecting threats is a highly complex cognitive task. Broadly conceptualized, it involves perceiving the time and space – the decision environment – in which a potential threat exists, evaluating the environment’s features, and generating hypotheses to explain those features holistically. These processes may change as a function of cognitive changes in the decision maker and perceived changes in the decision environment. For example, cues in an environment may be assigned different implicit threat values. A shadowy window in a roadside structure may appear to pose a greater threat than does a well-lit window. How Soldiers generate hypotheses to explain the likelihood of a threat posed by windows in a structure may depend on diverse factors, including whether the Soldier perceived the windows, the assigned threat value of the windows, whether additional cues in the environment are threat-relevant or -irrelevant, and the Soldier’s previous experience with similar cues.

Cue informativeness (i.e., threat value) and Soldier experience are critical, interrelated variables that influence hypothesis generation in threat detection. Cues may not possess implicit threat values. Rather, Soldiers assign threat values to cues based on semantic knowledge of and experience with those cues. Thus, if Soldiers do not recognize a cue, they may not be able to assign it a threat value. Consequently, they may erroneously fail to identify that cue as critical in assessing threat risk. Soldiers with limited experience in an operational environment may recognize fewer critical threat-relevant cues than would Soldiers with vast experience in the same environment. As a result, inexperienced Soldiers may be more likely than experienced Soldiers to evaluate a larger set of mixed-value cues (vs. a small set of high-value cues), and therefore rely on a less efficient process to generate hypotheses. Indeed, in a previous study of the effects of experience and uncertainty on hypothesis generation and evaluation, experienced and inexperienced Soldiers differed in some aspects of hypothesis generation, including how they identified and prioritized critical threat cues (Leins et al., 2013). In that study, Soldiers read scenarios accompanied by a static image depicting the scenario. For each scenario, Soldiers generated an initial hypothesis, investigated details added to the scenario, and then revised their initial hypotheses if necessary. In general, Soldiers revised their initial hypotheses in few instances, but when they did, they tended to revise hypotheses in which they were less confident. This was especially true of experienced Soldiers. When uncertain about their initial assessments, experienced Soldiers took advantage of the opportunity to explore more information and enhance their certainty. In that study, the influence of the decision environment and experience on aspects of hypothesis generation and evaluation was examined, such processes were assumed to be effortful. The efficiency or quality of hypothesis generation that is less effortful or deliberate was not explored. Through the research reported here, we sought to build on the previous study and explore how the decision environment impacts the efficiency and quality of Soldiers’ hypothesis generation processes. The current research explored two possible influences of hypothesis generation in threat detection: the use of heuristics and individual versus collective decision-making.

Heuristics in Hypothesis Generation

If decision makers can reduce the amount of information required to generate a hypothesis, they can generate hypotheses more efficiently. One reductive strategy for generating a hypothesis efficiently is to invoke heuristics. Heuristics reduce the cognitive effort necessary

to make a decision (e.g., see Gigerenzer & Gaissmaier, 2011; Kahneman & Frederick, 2002; Shah & Oppenheimer, 2008). For this research, the definition of a heuristic provided by Gigerenzer and Gaissmaier (2011) was used: “a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods” (p. 454). Accordingly, heuristics are frugal strategies because they reduce the number of cues used to generate a hypothesis. They are often quick because they employ stop rules: Once a cue passes a threshold of relevance, the decision maker stops searching for additional information and generates a hypothesis. Thus, how much time and information Soldiers used to generate hypotheses in different contexts was measured.

Threat detection in an operational environment is often characterized by knowledge and time constraints. Soldiers are not always familiar with the environments they operate in, nor do they often have time to deliberate upon the meaning of environmental cues or the best course of action in an uncertain situation. Hence, detecting threats in an operational environment is ideal for using heuristics to guide decision-making (e.g., see Rieskamp and Hoffrage, 1999). When the cost of slow, effortful deliberation appears prohibitive (e.g., life threatening), it will be sacrificed for quicker, more efficient decision processes.

Image theory (Beach, 1990) provides a useful context for considering how threat cues in the environment may activate heuristics and invoke stop rules in a deliberative decision-making process. According to image theory, emerging cues can alter perceptions of the current environment. These cues can be discrete cues or patterns of cues that are positive, neutral, or negative, unexpected or expected, and internal or external (Lee, Mitchell, Wise, & Fireman, 1996; Lee, Mitchell, Holtom, McDaniel, & Hill, 1999). They are salient enough to cause shifts in how a person judges an environment. In the threat detection context, a Soldier exposed to these cues may re-evaluate the perceived threat level of an environment. For example, a Soldier who is familiar with an environment may assess the situation as posing a relatively low threat risk because all its elements appear typical of low-threat conditions (e.g., a village marketplace is bustling with activity). However, when one or more elements appear atypical (e.g., the usually busy marketplace is oddly void of people and activity) the Soldier may elevate the threat assessment to a higher risk level.

As can be gleaned from the marketplace example, the role of memory appears to be critical in generating threat-relevant hypotheses. Thomas, Dougherty, Sprenger, and Harbison (2008) identified several critical relationships between environmental cues, memory, and hypothesis generation. They found that data extracted from the environment serve as cues that activate details in episodic and/or semantic memory (i.e., schemas). When environmental cues correspond strongly with a schema, they tend to elicit judgments in which decision makers are highly confident. Consequently, decision makers forego exploring other cues and alternative judgments and they generate a hypothesis based on the environment’s match with the schema (e.g., see Gettys, Manning, Mehle, & Fisher, 1980). Thus, in the example of the typical versus atypical marketplace activity, different patterns of cues may correspond with different schemas representing different threat levels. Strong correspondence with a particular schema may lead to a quick hypothesis. This suggests that the order in which decision makers perceive environmental cues may influence the speed with which they generate hypotheses.

Multiple researchers have demonstrated that decision makers often evaluate environmental cues sequentially (e.g., Gigerenzer & Goldstein, 1996; Pachur & Marinello, 2013;

Garcia-Retamero & Dhimi, 2009). When applying the filter of a recognition-based heuristic to sequential evaluation, decision makers will generate hypotheses at the rate at which they recognize cues. For example, Pachur and Marinello (2013) found that U.S. Customs Enforcement Officers screening for smugglers used a one-reason decision-making heuristic in which they evaluated cues sequentially and stopped after finding a single, sufficient cue. The sooner they perceived a sufficient cue, the sooner they generated a hypothesis. Contrast the use of recognition-based heuristics with weighted-additive strategies in which decision makers evaluate the value of multiple cues before generating a hypothesis. Because weighted-additive strategies require evaluating many cues, hypothesis generation should take longer than when using one-cue, recognition-based heuristics. However, decision makers using recognition-based heuristics can still suffer delays if they fail to perceive a sufficiently informative cue before perceiving all or many other cues.

In addition to the influence of cue order on heuristic usage, decision makers' reliance on heuristics may vary with the domain or decision environment. Decision makers are more likely to engage in heuristic use when decision tasks are familiar and decision makers have experience in the domain (Hammond, 1988). By contrast, if decision makers do not have domain-relevant information and experiences stored in long-term memory, they may not generate hypotheses as automatically or quickly (for discussions of the interaction of expertise and domain familiarity, see Shanteau, 1992a, 1992b). Given Soldiers' training and experience, detecting threats should be a relatively familiar decision task, and one that should induce Soldiers to use heuristics. However, other features of the decision task may facilitate or hinder the use of heuristics. Researchers have noted that decision makers are better able to engage in cue retrieval when tasks have inherent limitations (e.g., chess) versus tasks with unconstrained boundaries (e.g., diagnosing medical conditions; Norman, Brooks, & Allen, 1989). Threat detection, like medical diagnosis, is a task with unconstrained boundaries. The threat detection decision environment can change in an infinite number of ways. Given such volatility and potential variance in threat decision environments, Soldiers may have difficulty using recognition-based heuristics to assess situations. To disentangle the effects of unbounded decision tasks from domain familiarity on heuristic use, the current research tested the amount and type of information Soldiers viewed before generating hypotheses in two types of unbounded decision tasks: one familiar (threat detection) and one unfamiliar (medical diagnosis).

Familiarity is not the only environmental factor that influences heuristic use. The time available for decisions may also influence whether and to what extent decision makers use heuristics when generating hypotheses. Consider that heuristics can operate as stop rules (as well as search rules and decision rules). When an environmental cue passes some threshold of informativeness (i.e., it has a recognizable diagnostic value for the decision task), it activates an associated hypothesis and the decision maker can stop searching for and evaluating environmental cues. However, when no cue passes the threshold of informativeness, the decision maker may continue searching the environment. This process becomes problematic when decision makers do not have an infinite amount of time with which to search. In this case, they can proceed in one of two ways: delay a decision because they were unable to generate a suitable hypothesis or adjust the threshold for identifying cues as reasonably informative and use a less informative cue (or set of cues) to generate a hypothesis. Findings reported by Thomas, Dougherty, Sprenger, and Harbison (2008) suggest that decision makers under time constraints may indeed adjust their criteria for generating hypotheses. They found that as the time available

to make a decision decreased, so did the number of plausible hypotheses generated by decision makers. Adjusting the threshold of informativeness may result in generating suboptimal hypotheses, but in some instances decisions based on suboptimal hypotheses may be preferred to failing to make a decision. To determine whether Soldiers would adjust their thresholds for cue informativeness in threat detection, the current research incorporated perceived time constraints in some of the hypothesis generation tasks.

The current research explored whether Soldiers employ heuristics when generating threat hypotheses, by testing whether they employ cue-activated stop rules. Across two experiments, the number and type of cues were measured Soldiers evaluated before generating a hypothesis. Experiment 1 tested individual Soldiers. Experiment 2 tested a different sample of Soldiers participating in groups of 3-4.

Experiment 1: Exploring Hypothesis Generation at the Individual Level

Design

Experiment 1 used a 2 x 2 x 2 (cue-order [high-value first vs low-value first] x familiarity [familiar vs. unfamiliar decision environment] x time pressure [low vs. high]) within-subjects design. Participants worked individually at laptop computers to engage in hypothesis generation tasks across experimental conditions. The cue presentation order, the familiarity of the decision environment/task, and the time pressure associated with each decision task were manipulated. To determine heuristic influence, the number of images viewed before Soldiers reported a change in their hypothesis was measured.

Independent Measures. Three variables across scenarios were manipulated: order of cue values, familiarity of decision environment, and time pressure. Both cue order and familiarity were crossed with time pressure. Cue order was nested within familiarity.

Order of cues introduced. All scenarios presented incoming cues. Red arrows flashed onscreen identifying the location of a new piece of information. Each incoming cue possessed either a high or low value of informativeness. A cue's value corresponded to how strongly it associated with a potential scenario status. A high-value cue in a threat detection context correlated strongly with a potential high threat risk according to subject matter experts (SME). A high-value cue in a medical diagnosis context correlated strongly with a high risk of a particular disease according to SMEs. In half of the scenarios, high-value cues appeared in position two in the cue sequence, whereas in the other half of the scenarios, high-value cues appeared in position three. Thus, the two orders of added cues were:

- Low-value cue early: Initial image → low-value cue added (position one) → low-value cue added (position two) → high-value cue added (position three).
- High-value cue early: Initial image → low-value cue added (position one) → high-value cue added (position two) → low-value cue added (position three).

Cue presentation order was crossed with time pressure, and nested cue order within familiarity.

Familiarity of decision environment. Half of the scenarios involved a familiar decision context, whereas the other half involved an unfamiliar decision context. Participants received general guidance on how to respond in these contexts. Familiar scenarios involved assessing the threat risk of proceeding down urban and rural roads or paths, or investigating structures in an operational environment. They were told to consider three levels of threat risk when assessing each scenario relevant to an operational environment: green (low), amber (moderate), and red (high). Unfamiliar contexts included decisions associated with diagnosing medical conditions in individuals in a hospital triage environment. Participants considered whether they would admit each patient to an emergency room that could not treat non-emergency cases or divert the patient to non-urgent care. Each participant responded to six familiar context scenarios and six unfamiliar context scenarios.

Time pressure. Each participant completed six scenarios with low time pressure and six scenarios with high time pressure. Each set of high time pressure scenarios included instructions indicating that participants had only two minutes to complete all 12 scenarios. In actuality, they had to complete only six scenarios in this set. A visual timer accompanied high time pressure scenarios. This timer depleted as time elapsed, so participants could monitor their status. Onscreen instructions notified participants prior to time pressure trials that (a) the timer paused while they typed responses and (b) they could stop a trial, and pause the timer at any time by pressing a mouse button and reporting a change to their assessment. Hence, participants could conserve time by registering assessments early in any time-pressured trial. The purpose of the high time pressure was to determine whether Soldiers would make decisions based on suboptimal cue values (i.e., low-value cues) when they may reasonably anticipate having access only to those cues. Time pressure was crossed with context familiarity and cue order.

Dependent Measures. To determine whether heuristics influence hypothesis generation, we recorded whether participants made a decision and how many images they viewed before making their decision. Participants viewed images until they signaled a change in hypothesis, at which time no additional images appeared. The last possible image viewed for any trial was the original image plus three added cues.¹ This measure allowed us to identify the cue or cues currently available to participants when they stopped a trial. The level of threat identified by participants in each hypothesis was coded and their level of confidence in their hypotheses was also measured.

Experimental Hypotheses. It was predicted that participants would engage heuristics and generate hypotheses quicker when experiencing (a) high-value cues early in the cue order, (b) familiar decision environments/tasks and (c) high time pressure. It also predicted cue order to interact with time pressure. We predicted that low-value cues presented early would trigger quick hypotheses when under time pressure, but would not trigger quick hypotheses when no time pressure exists. It was not expected high-value cues to trigger hypotheses differentially as a function of time pressure. Last, the assumption was made that high time pressure and unfamiliar decision environments paired with late-appearing high-value cues would yield suboptimal timing

¹ We measured response timing according to the image onscreen when a participant stopped the trial or when the trial ended naturally. There were seven possible stop points according to seven distinct images: (1) the original stimulus image, (2) Cue 1 with indicator arrows (e.g., see Appendix A, Image 4), (3) Cue 1 without indicator arrows (e.g., see Appendix A, Image 5), (4) Cue 2 with indicator arrows, (5) Cue 2 without indicator arrows, (6) Cue 3 with indicator arrows, and (7) Cue 3 without indicator arrows.

of changes in hypotheses. That is, high time pressure and unfamiliar environments would lead participants to stopping scenarios prior to the appearance of the high-value cue when that cue appeared late in the scenario.

Method

Participants. Thirty-three Soldiers were tested with a mean age of 24.36 years, a mean time in service of 4.03 years, and a mean time in their current rank of 1.49 years (see Table 1 for additional demographic data). No participants reported familiarity with medical triage or diagnosis procedures beyond basic first aid.

Table 1

Experiment 1 Sample Demographics

		n	%
Current Rank	E-3:	7	21
	E-4:	15	45
	E-5:	11	33
Number of reported training courses aiding threat detection ability	0:	7	21
	1:	11	33
	2+:	15	45
Deployed	Yes:	22	67
	No:	11	33
Of participants who deployed:		n	%
Number of deployments	1:	12	55
	2:	9	41
	3:	1	5
Number of times “outside the wire”	Never:	7	32
	< 1/month:	1	5
	1/month:	1	5
	> 1/month:	2	9
	1/week:	1	5
	> 1/week:	6	27
	Every day:	4	18

Note. Current military occupational specialty code (MOS) reported by participants (n participants in parentheses): 11b (15), 11c (1), 14g (1), 15p (2), 19k (1), 25s (1), 29e (1), 35f (1), 68w (1), 88m (3), 91b (3), 92a (1), 92f (1), 92w (1).

Materials. Participants interacted with laptop computers running the Psychology Experiment Building Language (PEBL; Mueller & Piper, 2014) application. The PEBL application presented all experimental stimuli and recorded all participant responses in the form of button presses and keyed text. Visual stimuli included scenarios presenting decision tasks and environments.

Scenarios. Participants interacted with 12 scenarios. Six scenarios presented threat detection (familiar) tasks; the other six presented medical diagnosis (unfamiliar) tasks. Each scenario included a short description accompanied by a static image (see Appendix A for examples). Each description presented the scenario context and decision requirements (e.g., “Your squad is working through a village, classifying routes. Your squad leader has asked you to assess the threat level of this part of the route”). The description remained onscreen for 12 seconds, which was adequate time to read all the text. Following the description, participants were instructed to assess the scenario (“What is the threat level here?”) and provide a confidence rating. The image remained onscreen throughout reporting. After participants entered their assessment and confidence rating, a new cue was added to the image every six seconds until three new cues had been added or until the participants stopped the trial to indicate a change to their assessment. A six-second presentation time was chosen to allow participants adequate time to view new details, but also to advance each trial at a practical pace. Red indicator arrows accompanied all newly added cues to draw participants’ attention to this incoming information. The arrows disappeared after one second and the new cue remained onscreen. The cues added over time varied in informativeness. Subject matter experts were consulted to determine cue informativeness values. Low informativeness values were assigned to cues that military SMEs identified as low priority threats in the threat detection context and that medical SMEs identified as symptoms unspecific to a particular syndrome or illness in the medical diagnosis context. By contrast, high informativeness values were assigned to cues identified as high priority threats or identified as symptoms highly specific to a particular syndrome or illness. The first cue added to each scenario image was always a low-value cue. The second and third cues varied randomly in informativeness (high vs. low), with only one high-value cue added to each scenario and counterbalanced in its position (second vs. third) across trials. Thus, each scenario contained one high-value cue presented in either the second or third order position. After a participant completed a scenario and indicated readiness to proceed (via button press), PEBL loaded the next scenario. This process continued until the participant completed all 12 scenarios. Experimenters predetermined the scenario and condition order for a fixed sample of participants. The PEBL ran a separate script for each order according to a subject identification number keyed into the program at the start of a session. Participants were assigned randomly to subject identification numbers.

Demographic questionnaire and decision-making scales. After completing all 12 hypothesis generation scenarios, participants completed a demographic questionnaire and two decision-making scales (see Appendix 2). The demographic questionnaire included questions about relevant military experience. The first decision-making scale participants completed was the Decision-Making Style Scale (Scott & Bruce, 1995). This scale measures five types of decision-making style: rational, intuitive, dependent, avoidant, and spontaneous. According to Scott and Bruce (1995), rational decision makers search for and logically evaluate alternative hypotheses. Intuitive decision makers rely on hunches and feelings. Dependent decision makers search for advice and direction from others. Avoidant decision makers attempt to abstain from making decisions. Spontaneous decision makers possess a sense of immediacy and a desire to expedite the decision process. The second scale participants completed was a shortened Need for Cognitive Closure (NFCC) Scale (Roets & Van Hiel, 2011; for the original scale, see Webster & Kruglanski, 1994). This scale measured participants’ dispositional desire to obtain answers on a given topic. Individuals high in NFCC prefer order, structure, and predictability versus disorder. They also possess a sense of urgency to reach swift decisions. Thus, there was

interest in exploring whether any of these decision-making styles correlated with how much information participants would evaluate before generating new hypotheses.

Procedure. After consenting to participate, participants received general experimental instructions on how to interact with the PEBL application. They were told that they would provide an initial assessment of a scenario and then watch as that scenario changed over time. They were told that they should press the mouse button to stop a trial when they saw information that changed their initial assessment. They would enter their changed assessment, report their confidence in the change, and then move on to the next scenario. After these instructions, participants completed a set of practice trials to become familiar with the procedure. Practice trials asked for hypotheses about one familiar and one unfamiliar context. Participants were then introduced to the timer used in high time pressure trials. Participants then received general task instructions according to their experimental condition (e.g., participants were told that they had either two minutes or no time limit to complete the next 12 scenarios). Specific scenario instructions paired with relevant images followed the general instructions. For each scenario, participants generated an initial hypothesis relevant to the specific instructions (e.g., What is the threat level? Admit or divert this patient?). After reporting an initial assessment, participants reported their confidence in their assessment on a scale of 0-100. Following this entry, participants indicated their readiness to proceed and then viewed the incoming information (high- or low-value cues) for that scenario. They pressed the mouse button to stop a trial when their assessment changed. If they did not press the mouse button (either because their assessment did not change or they did not want to report a change), the trial concluded six seconds after the third cue appeared. At this time, participants received a prompt to report their assessment and confidence. Thus, Soldiers reported an initial and a second hypothesis for each scenario. After reporting their second hypothesis, participants indicated their readiness to proceed and continued on the next scenario. After completing 12 scenarios, participants completed the demographic questionnaire, the Decision-Making Style Scale and the NFCC Scale.

Scoring. The PEBL application output all data into Microsoft Excel files. It populated each file with a subject identification, data corresponding to each trial completed (i.e., the trial condition, the initial hypothesis and its associated confidence rating, the second hypothesis and its associated confidence rating, and the serial order position of the image onscreen when the trial concluded), and all responses to demographic and scale questions.

Two coders scored each hypothesis according to the threat level identified by the participant. Coders scored hypotheses on a scale of 0-2 (0 = low or minimal threat/urgency, 1 = moderate threat/urgency, 2 = high threat/urgency; see Appendix C for examples of scored hypotheses). Coders scored 257 hypotheses and reached good interrater reliability, $Kappa = .82$. Experimenters then calculated the change in identified threat level for each scenario by subtracting the value of the initial hypothesis from the value of the second hypothesis. Thus, a positive change in threat level score reflected an increase in perceived threat.

Soldiers' second hypotheses was scored for the quality of their timing. The hypotheses were scored as "optimal" if they occurred in conjunction with the appearance of the high-value cue, but "suboptimal" if they occurred prior to the high-value cue or if they did not occur until the end of a trial in which the high-value cue appeared early. Changes were compared in threat value across optimal versus suboptimal responses.

Data was scored for eight of the experimental scenarios. Data was removed corresponding to four scenarios (two familiar and two unfamiliar) that we designed as “dud” scenarios. Dud scenarios presented no high-value cues as incoming information. Duds were implemented to prevent participants from learning that a high-value cue would appear in every scenario and hence anticipating the value of incoming cues based on a pattern of cue values in completed scenarios. Thus, scenarios interleaved with no high-value cues among scenarios in which high-value cues appeared in different serial positions. Cue order was counterbalanced. After removing dud scenarios, threat level scores, the number of images viewed, and confidence values were averaged for the remaining scenarios within a given condition. Data of four participants was removed who were either outliers or whose data were insufficient to allow scoring (e.g., they failed to record hypotheses or confidence ratings). The following analyses included the remaining 29 participants.

Results

All hypothesis threat-level, latency, and confidence data were analyzed using paired-samples *t*-tests and analyses of variance (ANOVA) at $\alpha = .05$.²

Initial hypothesis threat-level. Across all scenarios ($N = 232$), Soldiers tended to report low to moderate initial threat levels ($M = 0.77$, $SD = 0.81$). Soldiers reported low or minimal threat levels in 47% of all scenarios, moderate threat levels in 29% of scenarios, and high threat levels in 24% of scenarios. Whether a Soldier had deployed did not influence the initial threat level reported for either the familiar or unfamiliar scenarios, both t values < 0.16 , both p values $> .87$.

Changes in hypothesis threat-level. Across all scenarios, Soldiers tended to report increases in threat levels over time ($M = 0.56$, $SD = 0.99$). Soldiers reported an increase in threat level in 48% of all scenarios, no change in threat level in 42% of scenarios, and a decrease in threat level in 10% of all scenarios. In 19% of scenarios initially reported as a low or minimal threat, participants did not report changes in threat level over time. In 45% of scenarios initially reported as a moderate threat, participants did not report changes in threat level over time. In 67% of scenarios initially reported as a high threat, participants did not report changes in threat level over time.

Cue order. A paired-samples *t*-test revealed no difference in the change to a reported threat level when the high-value cue appeared early ($M = 0.48$, $SD = 0.66$) versus when it appeared late ($M = 0.65$, $SD = 0.64$), $t(27) = 0.98$, $p = .33$, $d = 0.19$.

Cue order and familiarity were nested variables and therefore could not interact. Separate independent-samples *t*-tests were used to describe each nested relationship (see Table 2 for descriptive statistics). Within the familiar scenarios, cue order did not influence changes in participants' reported threat level, $t(26) = 0.21$, $p = .84$. Similarly, within the unfamiliar scenarios, cue order did not influence changes in reported threat level, $t(26) = 0.50$, $p = .62$.

² Depending on whether the independent variables (IV) were treated as within, between, or nested affected the degrees of freedom reported for some analyses.

Table 2

Threat Level Change: Descriptive Statistics for Cue Order Nested within Familiarity

		<i>M</i>	<i>SD</i>	N
Familiar	HV Cue Early	0.35	0.42	13
	HV Cue Late	0.32	0.35	15
	Total	0.33	0.37	28
Unfamiliar	HV Cue Early	0.73	0.87	15
	HV Cue Late	0.88	0.69	13
	Total	0.80	0.78	28

Note. HV = High-value.

Time pressure. A paired-samples t-test revealed that time pressure also influenced Soldiers' reported changes in threat level. Soldiers reported larger increases to perceived threat when experiencing no time pressure ($M = 0.67$, $SD = 0.51$) than when under time pressure ($M = 0.44$, $SD = 0.54$), $t(28) = 2.51$, $p = .018$, $d = 0.47$.

Time pressure \times familiarity interaction. A two-way repeated-measures ANOVA revealed no interaction effect of time pressure and familiarity on changes in reported threat level, $F(1, 28) = 0.07$, $p = .79$, $partial \eta^2 = .003$.

Time pressure \times cue order interaction. A two-way repeated measures ANOVA revealed no interaction effect of time pressure and cue order on changes in reported threat level, $F(1, 27) = 0.43$, $p = .52$, $partial \eta^2 = .015$.

Images viewed. Images viewed refers to the number of images viewed before participants signaled a change in their hypothesis (either by a participant's button press or after the final cue was introduced). This measure serves as a proxy for latency to respond. Across all scenarios, the number of images viewed correlated positively with the number of cues reported in hypothesis changes, $r = .196$, $p = .003$.

Cue order. A paired-samples t-test revealed no difference in images viewed when the high-value cue appeared early ($M = 5.34$, $SD = 1.34$) versus when it appeared late ($M = 5.75$, $SD = 1.40$), $t(27) = 1.07$, $p = .296$, $d = 0.20$.

Familiarity. A paired-samples t-test revealed no difference in images viewed when participants generated hypotheses in familiar contexts ($M = 5.54$, $SD = 1.33$) versus unfamiliar contexts ($M = 5.78$, $SD = 1.39$), $t(28) = 1.21$, $p = .237$, $d = 0.22$.

Independent-samples t-tests were conducted to explore the nested relationship between familiarity and cue order. In familiar scenarios, cue order did not influence the number of images viewed, $t(27) = -0.58$, $p = .57$. Similarly, in unfamiliar scenarios, cue order did not influence the number of images viewed, $t(26) = 0.30$, $p = .77$ (See Table 3 for descriptive statistics).

Table 3

Images Viewed: Descriptive Statistics for Cue Order Nested within Familiarity

		<i>M</i>	<i>SD</i>	N
Familiar	HV Cue Early	5.38	1.00	13
	HV Cue Late	5.68	1.62	15
	Total	5.54	1.35	28
Unfamiliar	HV Cue Early	5.67	1.60	15
	HV Cue Late	5.83	1.17	13
	Total	5.74	1.39	28

Notes. HV = High-value. The maximum possible number of images viewed was 7.

Time pressure. A paired-samples t-test revealed a main effect of time pressure on images viewed. Participants tended to report changes in their hypotheses sooner (i.e., after viewing fewer images) under time pressure ($M = 5.46$, $SD = 1.41$) versus no time pressure ($M = 5.87$, $SD = 1.28$), $t(28) = 2.22$, $p = .035$, $d = 0.41$.

Time pressure x familiarity interaction. A two-way repeated measures ANOVA revealed no interaction effect of time pressure and familiarity, $F(1, 28) = 0.13$, $p = .72$, *partial* $\eta^2 = .005$ (see Table 4 for descriptive statistics). However, the difference in latency between familiar-pressure conditions and unfamiliar-no pressure conditions was statistically significant, $t(28) = 2.56$, $p = .016$, $d = 0.48$.

Table 4

Images Viewed: Descriptive Statistics for Time Pressure x Familiarity

	<i>M</i>	<i>SD</i>	N
Familiar-No Pressure	5.77	1.42	29
Familiar-Pressure	5.38 ^a	1.53	29
Unfamiliar-No Pressure	5.91 ^a	1.45	29
Unfamiliar-Pressure	5.66	1.61	29

Note. ^a Denotes a statistically significant difference at $\alpha = .05$.

Time pressure x cue order interaction. A two-way repeated measures ANOVA revealed no interaction effect of time pressure and cue order on images viewed, $F(1, 27) = 0.65$, $p = .43$, *partial* $\eta^2 = .02$ (see Table 5 for descriptive statistics). However, a simple main effect was observed for time pressure when the high-value cue appeared late in the trial. In these scenarios, time pressure yielded earlier responses than did no time pressure, $t(27) = 2.96$, $p = .006$, $d = 0.57$. A similar difference was observed in scenarios in which the high-value cue appeared early, but this difference did not reach statistical significance, $t(28) = 1.95$, $p = .06$.

Table 5

Images Viewed: Descriptive Statistics for Time Pressure x Cue Order

	<i>M</i>	<i>SD</i>	<i>N</i>
No Pressure-HV Cue Early	5.79	1.39	28
No Pressure-HV Cue Late	5.98 ^a	1.48	28
Pressure-HV Cue Early	5.13	1.90	28
Pressure-HV Cue Late	5.04 ^a	1.79	28

Notes. ^a Denotes a statistically significant difference at $\alpha = .05$. HV = High-value.

Deployment history. A set of two-way mixed factorial ANOVAs revealed no interaction effect of deployment history (yes vs. no) and any primary independent variable (cue order, familiarity, or time pressure) on images viewed, all F values < 0.87 , all p values $> .36$. A subsequent set of ANOVAs revealed no interaction effect of number of times “outside the wire” and any primary independent variable on images viewed, all F values < 1.74 , all p values $> .19$.

Quality of timing. Quality of timing refers to whether participants stopped a scenario at an optimal or suboptimal time. Hypothesis changes that coincided with the appearance of the high-value cue were considered optimal (value = 1). Hypothesis changes that preceded a high-value cue or coincided with the natural conclusion of a scenario in which the high-value cue appeared early were considered suboptimal (value = 0). Overall, participants’ timing was optimal in 46% of scenarios and suboptimal in 54% of scenarios.

Cue order. A paired-samples t -test revealed an influence of cue order on the quality of timing. Participants were more likely to change their hypotheses at an optimal time when the high-value cue appeared late ($M = .58$, $SD = .35$) versus when it appeared early ($M = .32$, $SD = .35$), $t(27) = 2.73$, $p = .011$, $d = 0.53$.

Familiarity. A paired-samples t -test revealed no difference in the quality of timing when the scenario was familiar ($M = .53$, $SD = .38$) versus when it was unfamiliar ($M = .39$, $SD = .35$), $t(28) = 1.33$, $p = .193$, $d = 0.25$.

Two independent-samples t -tests revealed an influence of cue order nested within familiarity. When completing unfamiliar scenarios, participants were more likely to change their hypotheses at a more optimal time when the high-value cue appeared late versus early, $t(26) = 4.16$, $p < .001$, $d = 1.57$. When completing familiar scenarios, cue order did not influence the quality of timing, $t(26) = 1.56$, $p = .13$ (see Table 6 for descriptive statistics).

Table 6

Quality of Timing: Descriptive Statistics for Cue Order Nested in Familiarity

		<i>M</i>	<i>SD</i>	N
Familiar	HV Cue Early	.42	.34	13
	HV Cue Late	.63	.36	15
	Total	.54	.36	29
Unfamiliar	HV Cue Early	.20 ^a	.25	15
	HV Cue Late	.63 ^a	.30	13
	Total	.40	.35	28

Notes. ^a Denotes statistically significant difference at $\alpha = .05$. HV = High-value

Time pressure. A paired-samples t-test revealed no main effect of time pressure on the quality of timing. However, participants were more likely to change their hypotheses at an optimal time when under no perceived time pressure ($M = .50$, $SD = .28$) versus when under perceived time pressure ($M = .41$, $SD = .25$), $t(28) = 1.98$, $p = .057$, $d = 0.37$.

Time pressure x familiarity interaction. A two-way repeated measures ANOVA revealed no interaction effect of time pressure and familiarity on the quality of timing, $F(1, 28) = 0$, $MS = 0$, $p = 1.0$, $partial \eta^2 = 0$.

Time pressure x cue order interaction. A two-way repeated measures ANOVA revealed no interaction effect of time pressure and cue order on the quality of timing, $F(1, 27) = 0.13$, $MS = 0.01$, $p = .72$, $partial \eta^2 = .01$ (see Table 7 for descriptive statistics).

Table 7

Quality of Timing: Descriptive Statistics for Time Pressure x Cue Order

	<i>M</i>	<i>SD</i>	N
No Pressure-HV Cue Early	.34	.41	28
No Pressure-HV Cue Late	.63	.40	28
Pressure-HV Cue Early	.29	.35	28
Pressure-HV Cue Late	.54	.41	28

Note. HV = High-value.

Initial Confidence. Initial confidence refers to participants' ratings (0-100) of their confidence in their initial hypothesis. Initial confidence ratings were analyzed across familiarity and time pressure only. Because initial confidence was reported before cue order was manipulated in any trial, cue order should not influence initial confidence. Hence, the effect of cue order on initial confidence was not analyzed.³

Familiarity. A paired-samples t-test revealed a main effect of familiarity on confidence. Participants reported higher confidence when reporting hypotheses for familiar contexts ($M =$

³ To ensure cue order did not influence confidence, we analyzed its potential effect, which was null, $t(27) = 0.28$, $p = .78$, $d = 0.05$.

90.53, $SD = 12.06$) versus when reporting hypotheses for unfamiliar contexts ($M = 86.91$, $SD = 13.83$), $t(28) = 2.19$, $p = .037$, $d = 0.41$.

Time pressure. A paired-samples t-test revealed no difference in reported confidence when no time pressure existed ($M = 88.78$, $SD = 12.53$) versus when under time pressure ($M = 88.66$, $SD = 13.13$), $t(28) = 0.09$, $p = .932$, $d = 0.02$.

Familiarity x time pressure interaction. A two-way repeated measures ANOVA revealed no interaction effect of familiarity and time pressure on initial confidence, $F(1, 28) = 0.06$, $MS = 2.95$, $p = .81$, $partial \eta^2 = .002$ (see Table 8 for descriptive statistics).

Table 8

Confidence: Descriptive Statistics for Familiarity x Time Pressure

	<i>M</i>	<i>SD</i>	<i>N</i>
Familiar-No Pressure	90.76	11.93	29
Familiar-Pressure	90.31	13.21	29
Unfamiliar-No Pressure	86.81	15.55	29
Unfamiliar-Pressure	87.00	15.15	29

Change in confidence. The values reported in the following analyses represent the difference in confidence ratings between the reported confidence in the initial hypotheses and confidence in the second hypothesis. Changes in confidence did not correlate with quality of timing in any condition, all r values $< .28$, all p values $> .15$.

Cue order. A paired-samples t-test revealed no difference in changes in reported confidence when the high-value cue came early ($M = 0.86$, $SD = 3.21$) versus when it came late ($M = 1.04$, $SD = 7.89$), $t(27) = 0.11$, $p = .917$, $d = 0.02$.

Familiarity. A paired-samples t-test revealed no difference in changes in reported confidence when responding to familiar contexts ($M = 0.63$, $SD = 4.04$) versus when responding to unfamiliar contexts ($M = 1.29$, $SD = 7.31$), $t(28) = 0.39$, $p = .70$, $d = 0.07$.

Two independent-samples t-tests revealed no influence of cue order within familiar scenarios, $t(26) = 0.19$, $p = .85$, or within unfamiliar scenarios $t(26) = 0.06$, $p = .96$ (see Table 9 for descriptive statistics).

Table 9

Confidence Change: Descriptive Statistics for Cue Order Nested in Familiarity

		<i>M</i>	<i>SD</i>	N
Familiar	HV Cue Early	0.44	3.47	13
	HV Cue Late	0.75	4.72	15
	Total	0.61	4.11	28
Unfamiliar	HV Cue Early	1.22	3.04	15
	HV Cue Late	1.38	10.67	13
	Total	1.29	7.45	28

Note. HV = High-value.

Time pressure. A paired-samples t-test revealed no difference in changes in reported confidence when no time pressure existed ($M = 0.36$, $SD = 5.42$) versus when under time pressure ($M = 1.56$, $SD = 6.84$), $t(28) = 0.65$, $p = .52$, $d = 0.12$.

Time pressure x familiarity interaction. A two-way repeated measures ANOVA revealed no interaction effect of time pressure and familiarity on changes in confidence, $F(1, 28) = 0.03$, $MS = 2.49$, $p = .87$, $partial \eta^2 = .001$ (see Table 10 for descriptive statistics).

Table 10

Confidence Change: Descriptive Statistics for Time Pressure x Familiarity

	<i>M</i>	<i>SD</i>	N
Familiar-No Pressure	-0.16	5.76	29
Familiar-Pressure	1.41	6.47	29
Unfamiliar-No Pressure	0.88	10.51	29
Unfamiliar-Pressure	1.86	11.87	29

Time pressure x cue order interaction. A two-way repeated measures ANOVA revealed no interaction effect of time pressure and cue order on changes in confidence, $F(1, 27) = 0.19$, $MS = 16.90$, $p = .67$, $partial \eta^2 = .007$ (see Table 11 for descriptive statistics).

Table 11

Confidence Change: Descriptive Statistics for Time Pressure x Cue Order

	<i>M</i>	<i>SD</i>	N
No Pressure-HV Cue Early	0.80	3.00	28
No Pressure-HV Cue Late	0.21	11.37	28
Pressure-HV Cue Early	0.91	5.40	28
Pressure-HV Cue Late	1.88	12.59	28

Note. HV = High-value

Decision-Making Scales. Participants' scores on the NFCC Scale correlated with the number of images viewed under conditions of time pressure and early presentation of the high-

value cue (see Table 12). In both conditions, higher NFCC scores correlated with a greater number of images viewed. Linear regression analyses revealed that NFCC scores were not significant predictors of images viewed in these conditions. In addition, NFCC scores did not correlate with the quality of timing of hypothesis changes, all r values $< .15$, all p values $> .44$. No significant correlations were observed between Decision Making Style sub-scales and any outcome: All r values $< .21$, all p values $> .26$.

Table 12

Need for Cognitive Closure Scale Scores Association with Images Viewed

	Images Viewed			Correlation		Regression			
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>r</i>	<i>p</i>	<i>r</i> ²	<i>B</i>	<i>T</i>	<i>P</i>
Familiar	5.53	1.35	28	0.28	0.08	0.08	0.49	1.48	0.15
Unfamiliar	5.77	1.41	28	0.25	0.10	0.06	0.46	1.32	0.20
No Pressure	5.86	1.30	28	0.19	0.17	0.04	0.32	0.98	0.34
Pressure	5.44	1.43	28	0.34	0.04	0.12	0.63	1.84	0.08
HV Cue Early	5.54	1.35	28	0.33	0.04	0.11	0.58	1.79	0.09
HV Cue Late	5.73	1.43	27	0.19	0.17	0.04	0.36	0.99	0.33

Notes. NFCC $M = 4.10$, $SD = 0.76$. HV = High-value.

Discussion

In general, Soldiers tended to delay reporting changes in their hypotheses until evaluating all or nearly all of the information available in a scenario. This suggests they employed weighted-additive strategies for generating new hypotheses. Such strategies benefited Soldiers when high-value cues appeared late in scenarios. In these conditions, delaying a change in hypothesis associated with optimal timing. However, Soldiers appeared to change their strategies for generating hypotheses as a function of perceived time pressure. As the time allotted to complete a fixed number of scenarios decreased, Soldiers viewed fewer images before changing their hypotheses. This finding is not surprising, as one might expect decision makers to generate hypotheses for a set of problems more quickly when the time they have to do so decreases but the number of problems remains the same. One strategy decision makers can use to accelerate hypothesis generation is to adopt a satisficing heuristic. Satisficing is described as choosing a solution (or generating a hypothesis) that is not necessarily the best option, but may still be an effective option (Simon, 1957; see also Gigerenzer & Goldstein, 1996). As noted previously, this type of heuristic involves evaluating cues sequentially and stopping after finding a cue that surpasses a threshold of relevant informativeness (e.g., a criterion that helps categorize cues as high vs. low threat risk). Thus, given sequences of cues of varying informativeness, decision makers can lower their threshold of informativeness, allow a wider range of cue values to trigger a hypothesis, and settle on a candidate hypothesis after evaluating less information than would be required by a higher threshold of informativeness. Reducing this threshold, however, increases the risk of generating a suboptimal hypothesis when low-value cues precede high-value cues in the evaluation order. Soldiers experienced these conditions in Experiment 1. They received cues of varying informativeness, in random order, and generated hypotheses under time pressure. In response to this time pressure, Soldiers perhaps shifted their hypothesis generation strategy from one of a weighted-additive nature (evaluating each successive cue until no more

cues were available) to one of a satisficing nature. The observed effects of time pressure on both the number of images viewed and the quality of the timing of changes to hypotheses support this explanation. When under time pressure, Soldiers were more likely to change their hypotheses sooner and at suboptimal times compared to when under no time pressure.

However, no interaction of time pressure and cue order was observed. If Soldiers adopted a satisficing strategy and changed their hypotheses sooner, we would expect to see increases in the quality of timing for those hypothesis changes that coincided with the early appearance of high-value cues. Instead, quality of timing scores were *lowest* when the high-value cue appeared early in time-pressured scenarios. To help explain this, the timing of hypothesis changes under time-pressured scenarios were examined in which high-value cues appeared early. Soldiers stopped 45% of these scenarios early to report changes in their hypotheses. Forty-eight percent of those changes coincided with the appearance of the high-value cue (i.e., they were optimally timed), but 52% of those changes preceded the appearance of the high-value cue. In just over half of all time-pressured trials in which the high-value cue appeared early, Soldiers changed their hypotheses too soon. Indeed, Soldiers may have adopted a satisficing strategy and reacted to the first additional cue they perceived, which was always a low-value cue. In the context of Experiment 1, this yielded poorly timed changes in hypotheses.

The correlation between the number of images viewed and the number of cues offered in support of changed hypotheses also suggests that Soldiers may have sacrificed evaluating additional information to save time. As Soldiers viewed more images per scenario, they reported more cues in support of their hypothesis changes. When Soldiers registered changes sooner, they also reported fewer cues. Thus, when pressed for time, Soldiers may have changed their hypotheses sooner, in response to less information, and potentially in response to suboptimal information.

Ordinarily, accelerated hypothesis generation may not be detrimental. Indeed, expert decision makers likely engage in satisficing to generate hypotheses even when no time pressure exists, presumably because the strategy saves time and resources for subsequent hypothesis testing (Klein & Brezovic, 1986). Moreover, satisficing often leads to serviceable hypotheses. In the context of this experiment it is worth noting that, despite sometimes generating hypotheses based on partial information, Soldiers generally shifted their judgments to be less risk tolerant. Their changes were often in the direction of evaluating somewhat ambiguous stimuli as threat relevant. In 66% of scenarios initially assessed as having low or moderate threat levels, Soldiers reported increased threat levels over time. They generally shifted assessments of low or moderate risk levels to moderate or high risk levels, respectively. Thus, similarly engaging a satisficing heuristic in an operational environment may in fact serve to enhance Soldier safety.

A closer examination of hypothesis generation in scenarios with time pressure and early presentation of high-value cues revealed that some Soldiers did not engage heuristics to more quickly generate hypotheses. This subset of Soldiers tended to view as much information as possible before generating hypotheses and exhibited a cognitive disposition that differed from Soldiers who were more likely to engage a heuristic. In general, Soldiers who delayed generating hypotheses exhibited higher scores on the NFCC scale. Individuals high in NFCC prefer decision environments characterized by structure, certainty, and predictability. Soldiers high in NFCC may have been seeking to enhance the perceived certainty and predictability of experimental scenarios by maximizing the amount of information they evaluated for each

scenario. This finding may have general implications. Knowing which Soldiers may delay critical decisions, and when they may delay them, is important. As noted, delaying threat-relevant decisions may be more protective than endangering, but such an outcome depends critically on the nature of the context and decision. In many threat detection contexts, delaying decisions may be more inherently risky than making quick, suboptimal decisions. Unfortunately, NFCC values did not associate with the quality of timing of hypothesis changes in Experiment 1; hence, we cannot comment on the quality of hypothesis generation as a function of NFCC. Regardless, the NFCC measure appears relatively robust (Roets & Van Hiel, 2011; Webster & Kruglanski, 1994) and is easy to obtain. In future studies, researchers could use the NFCC scale to investigate the moderating impact of cognitive disposition on decision-making across different decision tasks and environments.

Surprisingly, familiarity did not influence the number of images viewed in Experiment 1. It was expected that familiar (threat relevant) decision tasks would elicit more heuristic usage than would unfamiliar (medical diagnosis) decision tasks. Soldiers' mean number of images viewed differed across decision tasks in the predicted direction, but the difference was not reliable. Perhaps Soldiers simply perceived no incentive to generate hypotheses more or less quickly across the familiarity of decision tasks. Whether the task was familiar or unfamiliar, Soldiers may have perceived no harm in evaluating all possible information. This explanation is plausible, as differences were observed in the number of images viewed when Soldiers perceived time pressure. Alternatively, Soldiers' initial perceptions of threat level may have precluded any potential influence of familiarity on heuristic usage. Given initial perceptions of neutral (or minimal) threat levels for all scenarios, it was expected that Soldiers would recognize and utilize high-value cues more efficiently in familiar scenarios than in unfamiliar scenarios. For example, given an initial perception of no threat in an operational environment and the addition of a threat-relevant cue to that environment, it would be expected for Soldiers to recognize the cue and adjust their perception of threat level accordingly (i.e., to something other than no threat). By contrast, given an initial perception of no urgency in a medical diagnosis task and the addition of a syndrome-specific (i.e., threat-relevant) cue, it was expected that Soldiers would be less likely to recognize the cue and adjust their perception of urgency. Hence, their familiarity with operational environments and associated threat-relevant cues would lead them to make recognition-based decisions more efficiently in those familiar contexts than in unfamiliar contexts such as medical diagnosis. However, in Experiment 1, Soldiers did not report neutral or minimal initial threat levels equally across familiar and unfamiliar scenarios. They reliably reported higher initial threat levels for familiar than for unfamiliar scenarios. Therefore, Soldiers had less room to adjust their threat level assessments of familiar scenarios over time. In these scenarios, they may have perceived additional cues as informative but not sufficiently informative to raise the perceived threat level higher than initially reported. In a scenario in which an added cue would have raised the threat level from minimal to moderate, if the initial threat level were already perceived to be moderate, there would be no need to stop the scenario and report a change. Consequently, familiar scenarios and unfamiliar scenarios yielded the same delay in reporting changes to hypotheses but by different mechanisms. Familiar scenarios yielded delays perhaps because initial threat levels were elevated and additional high-value cues, although likely recognized and evaluated appropriately, were not sufficient to raise threat levels any higher. By contrast, unfamiliar scenarios yielded delays perhaps because Soldiers did not recognize high-value cues as sufficient to raise urgency levels and instead they changed their hypotheses based on their evaluation of all additional cues (i.e., a weighted-additive strategy).

This implies that not only familiarity with a decision environment, but also initial impressions of that decision environment can influence whether heuristics guide hypothesis generation.

Finally, a restricted range of the number of viewable images may have reduced the influence of familiarity on hypothesis generation. Because the amount of information available in each scenario was fixed, regardless of context, Soldiers were necessarily limited in how long they could delay reporting hypotheses. If the range of information has been extended (and the time) available in each scenario, perhaps Soldiers would have continued evaluating added information in unfamiliar scenarios, but generated hypotheses more quickly in familiar scenarios. That Soldiers were more confident in their hypotheses pertaining to familiar scenarios versus unfamiliar scenarios may be informative. Perhaps Soldiers were more comfortable with the amount of information available in familiar scenarios than in unfamiliar scenarios. To elevate Soldiers' confidence in their hypotheses in unfamiliar scenarios would have required even more information, whereas increasing the amount of additional information in familiar scenarios may not have increased confidence similarly. Thus, without time pressure, Soldiers may choose to evaluate as much information as necessary to generate a hypothesis that surpasses a confidence threshold. In Experiment 1, Soldiers were forced to report a hypothesis regardless of their confidence in it. Further experiments are necessary to determine the effects of the amount of information and time available for generating hypotheses on confidence and heuristic usage across contexts.

The effects observed in Experiment 1 were of cue order, task familiarity, and time pressure on Soldiers generating hypotheses when working alone. However, hypothesis generation, particularly in threat detection, may often be a collective process. Soldiers work in groups to make decisions and solve problems. Hence, concurrent with Experiment 1, whether groups of Soldiers engaged heuristics when generating hypotheses was explored, and whether they did so differently than individuals.

Experiment 2: Exploring Hypothesis Generation at the Collective Level

Latencies consistent with heuristic usage among Soldiers generating hypotheses individually were observed. Whether they also engaged heuristics when generating hypotheses in groups was then explored. Reimer and Katsikopolous (2004) found that groups comprising a majority of individuals who use recognition-based heuristics made better decisions than groups comprising members who primarily used other decision-making strategies. Hence, it was expected that Soldiers continue to employ heuristics to generate hypotheses even when working in groups. However, the collective process likely appears different from the individual process.

As hypothesis generation shifts from an individual to a combined individual and collective process, it shifts from a primarily cognitive process to a social-cognitive process. Thus, group dynamics likely influence hypothesis generation and the use of heuristics. Indeed, merely placing people in groups affects their performance on myriad tasks. Social-cognitive phenomena such as social loafing (Latane, Williams, & Harkins, 1979) and diffusion of responsibility (Darley & Latane, 1968) suggest that as individuals shift from working alone to working in a group, their performance decreases. Latane, Williams, and Harkins (2008) suggest that these effects can be explained by social impact theory (see Latane, 1981). According to social impact theory, the social forces perceived by individuals acting alone decrease in

perceived strength as they become diffused across multiple group members. Thus, in a hypothesis generation task, individuals working alone may be motivated to generate a plausible hypothesis quickly because of perceived pressures to produce, whereas individuals working collectively may be less motivated to perform optimally (see also Gettys et al., 1980 for a discussion of the detrimental influence of groups on hypothesis generation).

Sorkin, Luan, and Itzkowitz (2004) conducted a study demonstrating the social impact of group decision making. They presented sets of cues to groups of up to 10 participants who decided whether individual cues represented a signal (an element that occurred in a previous display of nine elements) or noise (an element that had not occurred in a previous display). Each participant viewed the set of cues, made a decision, and reported a likelihood rating (0-100) that the identified cue was in fact a signal. The decisions were then presented to the group and a designated group member made a final group decision. In some conditions, participants gave initial responses and then the group was polled for a final vote, whereas in other conditions, group members were allowed to deliberate prior to a vote. Sorkin, Luan, and Itzkowitz reported that as group size increased, accuracy tended to increase, but efficiency decreased. However, they found that this decrease in efficiency was attributable to decreases in individual efficiency rather than group-driven efficiency. They suggested that social loafing might cause this effect. Similarly, it might be expected to observe variable input from Soldiers generating hypotheses collectively versus individually. However, such variable input may result not from deleterious phenomena, such as social loafing, but from institutionally derived social dynamics. For example, when hypothesis generation discussions include Soldiers of various ranks, the hierarchical structure within the Army may result in higher ranking Soldiers contributing disproportionately more to those discussions. This does not suggest that lower ranking Soldiers will withhold cognitive effort; it may simply reflect a learned position of deference.

Reimer and Hoffrage (2003) identified two group decision-making strategies relevant to Experiment 2: majority wins (consensus based on a group vote) and truth wins (consensus as a deferment to one member with knowledge of the solution). In familiar (threat detection) scenarios, it may be expected to see a majority-wins strategy when groups comprise Soldiers of the same rank and experience. In these scenarios, an equitable contribution across group members was expected, multiple 'pro' and 'con' arguments supporting group members' proffered hypotheses, and some negotiation before settling on a final hypothesis. By contrast, when groups comprise Soldiers of differing ranks or experience, it may be expected to observe a truth-wins strategy, as lower ranking (or less experienced) Soldiers may defer to higher ranking (or more experienced) Soldiers. Under these conditions, it was expected that fewer arguments and negotiations and less discussion would be observed prior to settling on a final hypothesis. In unfamiliar contexts, expected observations included a greater proportion of majority-wins strategies, and more equitable contribution, because threat detection-specific experience and rank may be less critical in generating good hypotheses. However, given the salience of hierarchy in the Army, truth-wins strategies may be employed liberally even in unfamiliar contexts.

In Experiment 2, whether groups and group dynamics would influence heuristic usage in hypothesis generation was explored as well as the type of group decision-making strategies groups might employ across contexts. Experiment 2 used the same general design and method as Experiment 1, but with two minor differences: Soldiers participating in Experiment 2 worked

collectively to generate hypotheses; and, experimenters recorded and scored Soldiers' discussions to examine group influence on hypothesis generation.

Design

Similar to Experiment 1, the familiarity of the decision environment/task, the time pressure associated with each decision task, and the cue presentation order was manipulated. Thus, Experiment 2 used a 2 x 2 x 2 (cue-order [high-value first vs. low-value first] x familiarity [familiar vs. unfamiliar] x time pressure [low vs. high]) within-subjects design.

Independent Measures. The same three variables as in Experiment 1 were manipulated: order of cue values, familiarity of decision environment, and time pressure. All variables were crossed and counterbalanced.

Dependent Measures. Using PEBL, initial group hypotheses and confidence ratings, the number of images viewed before reporting changes to hypotheses, and group confidence in consensus hypotheses were recorded. Video recordings of group hypothesis generation for group communication dynamics were coded. Then each group member's proportion of contribution to hypotheses generation discussions was calculated.

Experimental hypotheses. It was not predicted that fundamental cognitive processes would change as a function of individual versus group context. However, it was expected that groups might view more images before signaling changes in hypotheses, simply because of the added time required to convey individual perceptions and assessments to the group and to reach consensus on a changed assessment. Moreover, we expected social dynamics to influence the hypothesis generation process. In scenarios in which some group members explicitly demonstrate task knowledge, or have higher rank and greater experience, the group may defer to those group members when attempting to arrive at consensus. Such deference may yield less equitable contribution across participants. However, when no group member demonstrates sufficient task knowledge, or when rank is constant across group members, the group may arrive at consensus after input from a greater number of group members. Such collaboration may yield more equitable contribution across participants.

Method

Participants. Forty-four Soldiers were tested in groups of 3-4 (four groups of 3 members and eight groups of 4 members). Each group comprised squad members who were familiar with each other (mean time spent in unit = 19 months). Individual group members presented with a mean age of 23.45 years, a mean of 2.97 years in service, and a mean of 1.23 years in their current rank (see Table 13 for additional demographic data). Soldiers participated in pre-formed groups. Groups were randomly assigned to experimental conditions, counterbalancing the order of scenario presentation.

Table 13

Experiment 2 Sample Demographics

		N	%
Current Rank	E-1:	1	2
	E-2:	4	9
	E-3:	10	23
	E-4:	21	48
	E-5:	8	18
Number of reported training courses aiding threat detection ability	0:	6	14
	1:	16	36
	2+:	22	50
Deployed	Yes:	38	86
	No:	6	14
Of participants who deployed:		n	%
Number of deployments	1:	29	76
	2:	8	21
	3:	1	3
Number of times “outside the wire”	Never:	11	29
	< 1/month:	3	8
	1/month:	3	8
	> 1/month:	7	18
	1/week:	5	13
	> 1/week:	5	13
	Every day:	4	11

Note. Current MOS reported by participants (n participants in parentheses): 11b (15), 12b (2), 12n (5), 19d (12), 19k (10).

Materials. The same stimulus materials were used as in Experiment 1. To allow groups to view scenario images, the images were projected onto a white screen or white board. To capture group discussion, experimental sessions were recorded using a Panasonic digital camcorder, model HC-V270K. The camcorder was positioned between the projection screen and the participating group to record participants’ verbal and non-verbal behavior (e.g., nods of agreement). To capture indications of hypothesis changes that individual participants wished to signal silently, push-button signal devices were provided that connected to a panel of LED indicators. The indicator panel was positioned between the participants and the camcorder, so that the camcorder would record the indicators, but participants could not see them. This positioning was intended to allow participants to indicate changes without necessarily influencing other participants.

Procedure. Similar procedures in Experiment 1 were used, with some modifications. Instead of viewing stimuli individually, participants viewed stimuli in groups of three or four. One group member was randomly selected to interact with the PEBL application via laptop to record hypotheses and confidence ratings. Participants were seated or stood in a row to allow for

uniform viewing conditions. To ensure that group participants experienced similar viewing conditions as those experienced by individuals in Experiment 1, it was ensured that their viewing distance and projection dimensions allowed them to see the same details (e.g., no more than 23 feet from a 120-inch screen to offer the same viewing experience as a distance of no more than three feet from a 15-inch screen). After viewing an initial scenario image, participants discussed their assessments, reached a consensus, and then input their hypothesis into PEBL. They progressed through scenarios in the same way as participants in Experiment 1, but when group members viewed an image that triggered a change in their hypotheses, they could signal their individual change silently with their push-button device and/or verbally notify the group member operating the laptop computer to stop the scenario. Participants then discussed their assessments, reached consensus, and recorded a second hypothesis and confidence rating. After completing 12 scenarios, participants worked individually, at separate laptop computers, to complete the demographic questionnaire, the Decision-Making Style scale, and the NFCC scale.

Scoring. Quantitative data were scored in the same way as Experiment 1. Qualitative data were scored to identify individual group members' contribution to hypothesis generation discussions. To score for group member contribution, we reviewed each group's video and coded each group members' substantive statements and relevant non-verbal behavior. Substantive statements included suggestions regarding threat level (e.g., "This is definitely amber") or severity of a patient's symptoms (e.g., "There is nothing here that requires urgent attention"). This also included scoring statements of agreement or disagreement and supporting arguments. Relevant non-verbal behavior included nodding to indicate agreement, head-shaking to indicate disagreement, and various hand and finger gestures to indicate agreement or numerical statements.⁴ Statements not scored as contribution included irrelevant sidebars, meta-discussion about grammar, syntax, or spelling, and repetitions for the purpose of dictating items already discussed and agreed upon. These scores were used to calculate proportions of contribution for each participant in each scenario and used these proportion scores to calculate the contribution variance for each group across scenarios and conditions. As group members contributed equitably to discussions, the group contribution variance decreased.

Results

Initial hypothesis threat-level. Across all scenarios ($N = 96$), groups of Soldiers tended to report low to moderate initial threat levels ($M = 0.82$, $SD = 0.83$). Groups reported low or minimal threat levels in 45% of all scenarios, moderate threat levels in 28% of scenarios, and high threat levels in 27% of scenarios.

Changes in hypothesis threat-level. Across all scenarios, groups tended to report increases in threat levels over time ($M = 0.73$, $SD = 0.89$). Groups reported an increase in threat level in 49% of all scenarios, no change in threat level in 50% of scenarios, and a decrease in threat level in 1% of all scenarios.

⁴ We scored contribution three ways: Through word count analysis of video transcripts, subjective scoring of video transcripts, and direct subjective scoring of video files. Scoring a sample of two groups (17% of data), we attained an error rate of roughly 3% across all three scoring methods. Hence, we scored the remaining groups directly from video data, as this was the most efficient method.

Cue order. A paired-samples t-test revealed no influence of cue order on changes in reported threat levels, $t(11) = 0.28, p = .78$. Two independent-samples t-tests revealed no influence of cue order nested within familiar scenarios, $t(10) = 0.09, p = .93$, or within unfamiliar scenarios, $t(10) = 0.27, p = .80$.

Time pressure. A paired-samples t-test revealed no influence of time pressure on the magnitude of groups' changes in reported threat levels, $t(11) = 1.15, p = .28$.

Time pressure x familiarity interaction. A two-way repeated measures ANOVA revealed an interaction effect of time pressure and familiarity on groups' reported changes in threat level, $F(1, 11) = 9.51, p = .01, \text{partial } \eta^2 = .46$ (see Table 14 for descriptive statistics). When completing familiar scenarios, groups reported larger increases in threat level when under no time pressure versus when under time pressure, whereas when completing unfamiliar scenarios, groups reported larger increases in threat level when under time pressure versus when under no time pressure.

Table 14

Threat Level Change: Descriptive Statistics for Time Pressure x Familiarity

	<i>M</i>	<i>SD</i>	<i>N</i>
Familiar-No Pressure	0.38	0.38	12
Familiar-Pressure	0.25	0.50	12
Unfamiliar-No Pressure	0.88 ^a	0.86	12
Unfamiliar-Pressure	1.42 ^a	0.60	12

Note. ^a No simple main effects reached statistical significance; however, the effect of time pressure in unfamiliar scenarios is worth noting, $t(11) = -2.00, p = .07, d = 0.58$.

Time pressure x cue order interaction. A two-way repeated measures ANOVA revealed no interaction effect of time pressure and cue order on changes to reported threat level, $F(1, 11) = 0.33, p = .58, \text{partial } \eta^2 = .03$.

Images viewed.

Cue order. A paired-samples t-test revealed a main effect of cue order on number of images viewed. Participants' viewed fewer images when the high-value cue appeared early ($M = 6.35, SD = 0.71$) versus when the high-value cue appeared late in the trial ($M = 6.75, SD = 0.45$), $t(11) = -3.50, p = .005, d = 1.01$.

Familiarity. A paired-samples t-test revealed no difference in images viewed in familiar contexts ($M = 6.51, DF = 0.70$) versus unfamiliar contexts ($M = 6.58, SD = 0.56$), $t(11) = -0.42, p = .68, d = 0.12$.

An independent samples t-test revealed an influence of cue order nested within familiar scenarios, $t(10) = 2.42, p = .04, d = 1.43$. When completing familiar scenarios, groups viewed fewer images when high-value cues appeared early versus when they appeared late. By contrast, cue order had no effect on the number of images viewed in unfamiliar contexts (see Table 15 for descriptive statistics).

Table 15

Images Viewed: Descriptive Statistics for Cue Order Nested in Familiarity

		<i>M</i>	<i>SD</i>	<i>N</i>
Familiar	HV Cue Early	6.03 ^a	0.87	5
	HV Cue Late	6.86 ^a	0.24	7
Unfamiliar	HV Cue Early	6.57	0.53	7
	HV Cue Late	6.60	0.65	5

Notes. ^a Denotes statistically significant difference at $\alpha = .05$. HV = High-value.

Time pressure. A paired-samples t-test revealed no difference in images viewed when under no time pressure ($M = 6.69$, $DF = 0.56$) versus under time pressure ($M = 6.42$, $SD = 0.67$), $t(11) = 1.74$, $p = .11$, $d = 0.50$.

Time pressure x familiarity interaction. An ANOVA revealed no interaction effect of time pressure and familiarity on images viewed, $F(1, 11) = 0.06$, $p = .81$, $partial \eta^2 = .01$ (see Table 16 for descriptive statistics).

Table 16

Images Viewed: Descriptive Statistics for Time Pressure x Familiarity

	<i>M</i>	<i>SD</i>	<i>N</i>
Familiar-No Pressure	6.65	.59	12
Familiar-Pressure	6.42	.76	12
Unfamiliar-No Pressure	6.67	.65	12
Unfamiliar-Pressure	6.50	.67	12

Time pressure x cue order interaction. An ANOVA revealed no interaction effect of time pressure and cue order on images viewed, $F(1, 11) = 0.06$, $p = .80$, $partial \eta^2 = .01$ (see Table 17 for descriptive statistics).

Table 17

Images Viewed: Descriptive Statistics for Time Pressure x Cue Order

	<i>M</i>	<i>SD</i>	<i>N</i>
Pressure-HV Cue Early	6.08	1.24	12
Pressure-HV Cue Late	6.33	1.15	12
No Pressure-HV Cue Early	6.63	0.57	12
No Pressure-HV Cue Late	6.75	0.62	12

Note. HV = High-value.

Quality of timing. Hypothesis changes that coincided with the appearance of the high-value cue were considered optimal. Hypothesis changes that preceded a high-value cue or coincided with the natural conclusion of a scenario in which the high-value cue appeared early

were considered suboptimal. Overall, groups' timing was optimal in 57% of scenarios and suboptimal in 43% of scenarios.

Cue order. A paired-samples t-test revealed an influence of cue order on the quality of timing. Groups were more likely to change their hypotheses at an optimal time when the high-value cue appeared late ($M = .90$, $SD = .17$) versus when it appeared early ($M = .25$, $SD = .24$), $t(11) = 6.49$, $p < .001$, $d = 1.87$.

Familiarity. A paired-samples t-test revealed no difference in the quality of timing when the scenario was familiar ($M = .67$, $SD = .36$) versus when it was unfamiliar ($M = .48$, $SD = .41$), $t(11) = 0.89$, $p = .39$, $d = 0.26$.

Two independent-samples t-tests revealed influences of cue order nested within familiarity. When completing familiar scenarios, groups were more likely to change their hypotheses at a more optimal time when the high-value cue appeared late versus early, $t(9) = 6.84$, $p < .001$, $d = 4.16$. Similarly, when completing unfamiliar scenarios, groups changed their hypotheses at more optimal times when the high-value cue appeared late versus early, $t(10) = 4.33$, $p = .001$, $d = 2.41$ (see Table 18 for descriptive statistics).

Table 18

Quality of Timing: Descriptive Statistics for Cue Order Nested in Familiarity

		<i>M</i>	<i>SD</i>	<i>N</i>
Familiar	HV Cue Early	.30 ^a	.21	5
	HV Cue Late	.96 ^a	.10	6
	Total	.66	.37	11
Unfamiliar	HV Cue Early	.25 ^b	.27	6
	HV Cue Late	.85 ^b	.22	5
	Total	.52	.39	11

Notes. ^{a, b} Denote statistically significant differences at $\alpha = .05$. HV = High-value

Time pressure. A paired-samples t-test revealed no difference in the quality of timing when under no perceived time pressure ($M = .54$, $SD = .10$) versus when under perceived time pressure ($M = .60$, $SD = .20$), $t(11) = 1.00$, $p = .34$, $d = 0.29$.

Time pressure \times familiarity. A two-way repeated measures ANOVA revealed no interaction effect of time pressure and familiarity on the quality of timing, $F(1, 11) = 0.05$, $MS = 0.01$, $p = .82$, $partial \eta^2 = .01$.

Time pressure \times cue order. A two-way repeated measures ANOVA revealed no interaction effect of time pressure and cue order on the quality of timing, $F(1, 11) = 1.54$, $MS = 0.13$, $p = .24$, $partial \eta^2 = .12$.

Initial confidence.

Familiarity. A paired-samples t-test revealed no difference in reported confidence for hypothesis of familiar contexts ($M = 92.81$, $SD = 10.79$) versus unfamiliar contexts ($M = 94.35$, $SD = 8.77$), $t(11) = -1.16$, $p = .270$, $d = 0.33$.

Time pressure. A paired-samples t-test revealed no difference in reported confidence when no pressure existed ($M = 93.94$, $SD = 10.70$) versus when under time pressure ($M = 93.23$, $SD = 10.02$), $t(11) = 0.31$, $p = .765$, $d = 0.09$.

Familiarity x time pressure interaction. An ANOVA revealed no interaction effect of familiarity and time pressure on initial confidence, $F(1, 11) = 1.91$, $MS = 70.08$, $p = .19$, *partial* $\eta^2 = .15$ (see Table 19 for descriptive statistics).

Table 19

Confidence: Descriptive Statistics for Familiarity x Time Pressure

	<i>M</i>	<i>SD</i>	<i>N</i>
Familiar-No Pressure	94.38	11.78	12
Familiar-Pressure	91.25	13.03	12
Unfamiliar-No Pressure	93.50	10.22	12
Unfamiliar-Pressure	95.21	8.62	12

Change in confidence.

Cue order. A paired-samples t-test revealed no difference in changes to confidence when high-value cues appeared early ($M = -3.23$, $SD = 10.00$) versus when they appeared late ($M = 0.31$, $SD = 6.87$), $t(11) = -1.36$, $p = .20$, $d = 0.39$.

Familiarity. A paired-samples t-test revealed no difference in changes to confidence when responding to familiar contexts ($M = -0.31$, $SD = 10.06$) versus when responding to unfamiliar contexts ($M = -2.60$, $SD = 7.06$), $t(11) = 0.84$, $p = .42$, $d = 0.24$.

Two independent-samples t-tests revealed no influence of cue order nested in familiar scenarios, $t(10) = -1.08$, $p = .31$, or in unfamiliar scenarios, $t(10) = 0.04$, $p = .97$ (see Table 20 for descriptive statistics).

Table 20

Confidence Change: Descriptive Statistics for Cue Order Nested in Familiarity

		M	SD	N
Familiar	HV Cue Early	-4.00	11.94	5
	HV Cue Late	2.32	8.43	7
	Total	-0.31	10.06	12
Unfamiliar	HV Cue Early	-2.68	9.34	7
	HV Cue Late	-2.50	2.50	5
	Total	-2.60	7.06	12

Note. HV = High-value.

Time pressure. A paired-samples t-test revealed no difference in changes to confidence when no time pressure existed ($M = 0.31$, $SD = 7.01$) versus when under time pressure ($M = -3.23$, $SD = 10.23$), $t(11) = 1.26$, $p = .23$, $d = 0.36$.

Time pressure x familiarity interaction. An ANOVA revealed no interaction effect of familiarity and time pressure on change in reported confidence, $F(1, 11) = 0.13$, $MS = 18.75$, $p = .72$, $\text{partial } \eta^2 = .01$ (see Table 21 for descriptive statistics).

Table 21

Confidence Change: Descriptive Statistics for Time Pressure x Familiarity

	<i>M</i>	<i>SD</i>	N
Familiar-No Pressure	2.29	8.76	12
Familiar-Pressure	-2.92	15.84	12
Unfamiliar-No Pressure	-1.67	8.14	12
Unfamiliar-Pressure	-4.38	14.85	12

Time pressure x cue order interaction. An ANOVA revealed no interaction effect of time pressure and cue order on changes in reported confidence, $F(1, 11) = 0.98$, $MS = 102.08$, $p = .35$, $\text{partial } \eta^2 = .08$ (see Table 22 for descriptive statistics).

Table 22

Confidence Change: Descriptive Statistics for Time Pressure x Cue Order

	<i>M</i>	<i>SD</i>	N
No Pressure-HV Cue Early	0.00	7.46	12
No Pressure-HV Cue Late	0.63	9.78	12
Pressure-HV Cue Early	-6.46	17.79	12
Pressure-HV Cue Late	0.00	5.11	12

Note. HV = High-value.

Group-specific factors. The following analyses present associations among a group's time spent working together in the same unit, its group members' observed contributions to hypothesis discussions, and select outcome variables.

Individual participant contribution. A correlational analysis revealed that individual participant contribution correlated with rank, $r = .30$, $p = .048$. As a participants' rank increased, their contribution to hypothesis discussions increased. Participant contribution, however, did not vary as a function of the familiarity of the scenario, $t(43) = .03$, $p = .976$, nor was there an interaction effect of rank and familiarity on contribution, $F(4, 39) = 1.69$, $MS = .002$, $p = .17$, $partial \eta^2 = .148$. In addition, decision-making style and NFCC scores did not associate with individual contribution, all r values $< .23$, all p values $> .14$.

Contribution variance. The distribution of participant contribution (i.e., contribution variance) did not correlate with the number of images viewed in any condition, all r values $< .12$, all p values $> .70$. Similarly, contribution variance did not correlate with confidence in any condition, all r values $< .16$, all p values $> .61$. Contribution variance also was unaffected by the variance in ranks within groups, $r = .41$, $p = .19$.

Time spent working together. The amount of time a group spent working together correlated negatively with the distribution of participant contribution within the group, $r = -0.56$, $p = .06$. As the amount of time spent together increased, contribution variance decreased; thus, groups whose members spent more time together demonstrated more equitable participation across group members. No significant correlations were observed between time spent together and images viewed, all r values $< .21$, all p values $> .51$, or between the quality of timing of hypothesis changes, all r values < 1.6 , all p values $> .33$. No interaction effects of time spent together and any other variable on the number of images viewed were observed, all F values < 0.63 , all p values $> .44$. Similarly, no significant correlations were observed between time spent together and confidence, all r values $< .47$, all p values $> .12$, and no interaction effects on confidence were observed, all F values < 3.90 , all p values $> .07$.⁵

Individual versus group participants. Because participants interacted with the same stimuli across experiments, we were able to compare group hypothesis generation with individual hypothesis generation. The following analyses used participant status (individual vs. group unit) as an independent variable.

Independent-samples t-tests revealed differences in images viewed as a function of participant (individual versus group) status (see Table 23). In all conditions, groups of participants (Experiment 2) viewed more images than did individual participants (Experiment 1). These differences reached statistical significance in all conditions except when the high-value cue appeared early in scenarios. Participant status did not interact with any independent variable to influence the number of images viewed, all F values < 0.35 , all p values $> .56$, all $partial \eta^2$ values $< .01$. In addition, no three-way interactions reached statistical significance, all F values < 0.21 , all p values $> .65$, all $partial \eta^2$ values $< .01$.

⁵ These minimum values were obtained for the interaction effect of time spent together and time pressure on confidence, $F(1, 10) = 3.90$, $p = .08$, $MS = 98.93$, $partial \eta^2 = .281$. Regression analyses using time spent together as a predictor revealed no influence on confidence in either pressure condition, all r^2 values $< .22$, all p values $> .13$.

Table 23

Images Viewed: Effect of Individual versus Group

	Individual			Group			<i>t</i>	df	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	N	<i>M</i>	<i>SD</i>	N				
Familiar	5.54	1.33	29	6.51	0.70	12	-3.05	36.51	.004	0.82
Unfamiliar	5.78	1.39	29	6.58	0.56	12	-2.63	39.00	.012	0.66
No Pressure	5.87	1.28	29	6.69	0.56	12	-2.12	39.00	.040	0.73
Pressure	5.46	1.41	29	6.42	0.67	12	-2.95	38.08	.005	0.77
Early Cue	5.56	1.32	29	6.35	0.71	12	-1.94	39.00	.060	0.67
Late Cue	5.75	1.40	28	6.75	0.45	12	-3.38	36.41	.002	0.83

Independent-samples *t*-tests revealed differences in the quality of timing of hypothesis changes as a function of participant status (see Table 24). When under time pressure and when the high-value cue appeared late in a scenario, groups of participants reported changes to their hypotheses at more optimal times than did individual participants.

Table 24

Quality of Timing: Effect of Individual versus Group

	Individual			Group			<i>t</i>	df	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	N	<i>M</i>	<i>SD</i>	N				
Familiar	.53	.38	29	.67	.36	12	-1.10	39.00	.280	0.37
Unfamiliar	.39	.35	29	.48	.41	12	-0.72	39.00	.473	0.24
No Pressure	.50	.28	29	.54	.10	12	-0.71	38.52	.480	0.16
Pressure	.41	.25	29	.60	.20	12	-2.33	39.00	.025	0.80
Early Cue	.32	.34	29	.25	.24	12	0.74	29.19	.466	0.22
Late Cue	.58	.35	28	.90	.17	12	-3.87	37.32	.000	1.04

Independent-samples *t*-tests of the effect of participant status on confidence revealed differences only in unfamiliar contexts, in which individuals reported lower confidence ($M = 86.91$, $SD = 13.83$) than did their group counterparts ($M = 94.35$, $SD = 8.77$), $t(32) = -2.07$, $p = .047$, $d = 0.59$. For all other comparisons of confidence across participant status, all t values < 1.53 , all p values $> .13$. In addition, there was no difference in the magnitude of confidence change between individuals and groups, all t values < 1.76 , all p values $> .08$.

Discussion

Similar to Experiment 1, groups of Soldiers delayed reporting changes in their hypotheses until viewing all scenario images. In fact, groups of Soldiers delayed their changes even longer than did individual Soldiers in Experiment 1. These longer delays likely resulted from the added time group members needed to communicate and reach consensus. In contrast to Experiment 1, the order of cues influenced the number of images groups of Soldiers viewed

before changing hypotheses. When high-value cues appeared early, groups of Soldiers reported changes to their hypotheses sooner than when low-value cues appeared early. One possible explanation for this effect is that an increase in the number of participants viewing information and contributing to hypotheses increased the odds that at least one participant would recognize a high-value cue as informative and ask to stop the scenario. Unfortunately, we are unable to validate this explanation with empirical data. Although individual Soldiers could trigger light-emitting diode (LED) indicators to signal their perceptions prior to or concurrent with verbalizing their hypotheses, they did not signal these perceptions reliably. Some Soldiers in some groups used these devices, but too few used them to allow any conclusions on the timing of individual Soldiers' perceptions outside of their verbal contributions. Thus, comparisons of individual latencies in Experiment 1 with individual group member latencies in Experiment 2 are not possible. There is some evidence to suggest, however, that multiple decision makers working together can outperform individual decision makers on the same task. Laughlin and Shippy (1983) tested individuals versus groups on the ability to determine induction rules for categorizing playing cards as exemplars or non-exemplars. They found that groups solved more problems and offered larger proportions of plausible hypotheses than did individuals. Thus, it is reasonable to believe that multiple Soldiers generating hypotheses about operationally relevant scenarios might be quicker to generate plausible hypotheses than would individuals performing the same task. However, despite viewing fewer images when high-value cues appeared early, the quality of timing of groups' hypothesis changes did not also increase. In fact, the quality of timing of hypotheses changes was consistently higher when high-value cues appeared later in scenarios. This indicates that when groups stopped scenarios early to report hypothesis changes, their timing did not coincide with optimal information. A closer look at the timing of reported hypothesis changes suggests that although groups stopped scenarios earlier when high-value cues appeared earlier, they were likely to delay stopping those scenarios until after the high-value cue had been onscreen for several seconds and an additional low-value cue appeared. Stopping these trials between images four and five would have been optimal. On average, groups of Soldiers did not stop these trials until image six or later.

The familiarity of scenarios moderated the effect of cue order on the number of images viewed. When groups generated hypotheses in a threat detection context, they changed hypotheses sooner when high-value cues appeared earlier rather than later in the scenario. By contrast, when groups generated hypotheses in the medical diagnosis context, they were uninfluenced by the serial position of the high-value cue. Again, this effect may be a result of groups of Soldiers working together to solve recognition-based problems across two contexts that differentially confer advantages based on knowledge and experience. In the familiar context (threat detection), multiple Soldiers in the group likely possessed multiple relevant schemas allowing for rapid matches of environmental cues to patterns of cues in memory. By contrast, these same Soldiers likely possessed fewer relevant schemas for diagnosing medical conditions (unfamiliar context), and thus experienced delayed pattern matching or no pattern matching at all. Hence, the familiarity of the scenario apparently drove the observed effect of cue order. It is encouraging that groups of Soldiers worked together to generate relatively quick hypotheses in threat detection contexts; however, similar to the effect of cue order on the number of images viewed and the quality of timing of hypothesis changes, familiarity did not moderate the quality of timing of hypothesis changes. This, and the natural delay inherent in group decision-making (vs. individual decision-making), is possibly a result of the time required for group members to

communicate their perceptions and hypotheses. Further delays might result from disagreement among group members. Moreover, in Experiment 2, although multiple group members could request to stop a scenario, only one group member had access to the computer that could stop it. Thus, group members may have generated hypotheses coinciding with the appearance of high-value cues, but the time required to stop a scenario may have resulted in delaying the report until after additional low-value information appeared.

A critical question then arises: Who contributed critically to hypothesis generation discussions? In general, it was found that the proportion of contribution correlated with rank among group members. As rank increased, so did the proportion of contribution. This may reflect a tendency toward adopting a truth-wins strategy for reaching consensus (Reimer & Hoffrage, 2003), as it reflects a larger contribution from ostensibly the most knowledgeable group members and possible deference from less knowledgeable group members. One group of participants apparently adopted this strategy for generating hypotheses, as the highest ranking member essentially dictated hypotheses to the lower ranking members. In this case, the experimental session appeared more like a training session rather than a collaborative effort to generate hypotheses. By contrast, several groups appeared to adopt a majority-wins strategy for generating hypotheses (Reimer & Hoffrage, 2003). In these groups, each member contributed relatively equally. They exhibited discussion components that included presentation of hypotheses, supporting arguments, counter arguments, and negotiation prior to arriving at a consensus hypothesis. Surprisingly, however, the variance of rank within groups did not correlate with the variance of contribution within groups. Thus, in general, groups with members of varying ranks exhibited similar levels of equity in discussions as did groups with members of the same rank.

Even more surprising, the familiarity of scenarios did not influence contribution variance among groups, even among groups characterized by relatively high variance in ranks. Thus, hypothesis discussions were as equitable when some members may have been more experienced and knowledgeable (i.e., in threat detection scenarios) versus when all members should have possessed similar bases of knowledge (i.e., in medical diagnosis scenarios). This is surprising because one might expect scenarios with key features known to at least one group member to be more likely than scenarios with features unrecognized by group members to induce deference to group members with knowledge (Laughlin & Ellis, 1986). Thus, whereas in individual hypothesis generation the expectation was that heuristics would be activated by the decision space (task demands and environmental cues), in collective hypothesis generation it was expected that an additional influence of social dynamics on heuristic usage would be observed. Although the decision space did appear to influence heuristics in collective hypothesis generation, an influence of social dynamics on heuristic use was not observed.

The only group-member variable that correlated with the variance of contribution was the amount of time a group had been working together prior to the experiment. The longer a group spent working together, the lower the variance in group member contribution. Thus, groups whose members were more familiar with each other tended to exhibit more equitable discussions. This finding is encouraging, because it suggests that as Soldiers become more familiar with each other, they may be more willing to share their perceptions and interpretations with the group. Less encouraging, however, is that the factors that apparently influenced equitable contribution did not appear to influence how early in the image sequence Soldiers

generated hypotheses. Even though groups presented with varying amounts of time spent together, and may have exhibited different strategies for generating hypotheses, these conditions did not differentially affect the efficiency of hypotheses generation. One possible explanation for this mirrors the explanation for the null effect of familiarity in the first experiment: Perhaps a restricted range of scenario information and duration also artificially constrained any variability in the number of images viewed that may have occurred with longer scenarios. If Soldiers viewed longer scenarios, the gap between early and late responding may have differed as a function of different strategies for generating hypotheses. As noted in the discussion of Experiment 1, future studies should include scenarios of longer duration, to better examine how the amount of information required to change hypotheses might vary naturally across individual cognitive and group dynamic factors.

In contrast to individuals generating hypotheses, groups appeared to be uninfluenced by time pressure. Whereas time pressure induced individual Soldiers to engage satisficing heuristics, it apparently failed to induce the same strategy among groups. However, a similar trend was observed in the number of images viewed among groups as a result of time pressure – under time pressure, participants tended to make decisions quicker – but this trend did not reach statistical significance. This null effect may have resulted from one or both of two procedural factors: the restricted range cited earlier or a lack of power. In fact, the size of the effect of time pressure on the number of images viewed was larger for groups than for individuals; but because of limited power, its reliability is questionable.

General Discussion

Across experiments, the influences of cue order, decision task familiarity, and time pressure was observed on the number of images Soldiers viewed before reporting changes to their hypotheses. These findings suggest that Soldiers engaged different hypothesis generation strategies as a function of the context or decision space in which they operated. In some contexts, Soldiers used all of the allotted time and evaluated all of the possible information before reporting a new hypothesis. This reflects a strategy similar to a weighted-additive model of decision-making, in which a decision maker considers the values of all relevant attributes of a decision environment and their relative importance to the decision maker prior to settling on a hypothesis (see Payne, Bettman, & Johnson, 1993). Weighted-additive models of decision-making assume that decision makers are willing to make trade-offs when environmental cues have similar attribute values but vary in importance. For example, a suspicious vehicle and a moderate amount of ‘dead’ space in an operational environment may have moderate threat relevance, but one cue may be perceived as more important in the context of different decision tasks. Hence, a Soldier employing a weighted-additive strategy would evaluate both threats and weight more heavily, and base a hypothesis on, the cue or cues that appeared most important in the context of the operational question. Across experiments, Soldiers may have adopted weighted-additive strategies when they experienced no time pressure, when the decision context was unfamiliar, and when they received only low-value information early in the scenario. These conditions presented ambiguous information early – thus offering no obvious disparity in cue values – and they presented no consequence for waiting for more valuable information. In efforts to gain certainty in these scenarios, Soldiers simply waited for more information to evaluate.

Alternatively, Soldiers may have employed heuristics across all experimental conditions, but it was imperceptible until Soldiers encountered conditions that altered how they applied heuristics. In contrast to conditions that allowed ostensibly more time-consuming hypothesis generation strategies, increased time pressure, a familiar context, and early access to valuable information were conditions conducive to applying more efficient heuristics. In these cases, Soldiers may have adopted (or modified) heuristics that simplified the hypothesis generation process by limiting the amount of information they needed to process. Under time pressure, individual Soldiers appeared to employ satisficing heuristics. Such heuristics allow decision makers to evaluate cues one at a time and generate a hypothesis upon perceiving a sufficiently informative cue (Simon, 1957; see also Gigerenzer & Goldstein, 1996). If after one series of evaluations, no cue passes the threshold of informativeness, decision makers can lower the threshold and reapply the heuristic (see Payne, Bettman, & Johnson, 1993). Therefore, we cannot eliminate the possibility that Soldiers employed satisficing heuristics in ambiguous and unfamiliar contexts. Rather, Soldiers may have used them inefficiently, applying them iteratively and modifying them only when necessary. By contrast, Soldiers in Experiment 1 may have modified their heuristics a priori under certain conditions. This explains how Soldiers generated hypotheses sooner under time pressure. Time pressure combined with a delay in receiving critical information seemingly induced individual Soldiers to lower their criterion for judging the informativeness of environmental cues and to trigger new hypotheses.

It was also observed that familiar decision environments promoted quicker hypothesis generation among groups of Soldiers. These groups seemingly leveraged recognition-based heuristics in a way that individual Soldiers did not or could not. The effects that represent heuristic-based hypothesis generation suggest that under different conditions Soldiers may generate hypotheses at different rates of efficiency. What is unclear, however, is whether Soldiers use these more efficient processes to generate good hypotheses. There is some disagreement in the literature regarding whether heuristics, which by nature lead to ignoring information, also lead to good, accurate decisions (e.g., see Gigerenzer & Brighton, 2009) or to biased, faulty decisions (e.g., see Tversky, & Kahneman, 1974). In the experiments presented here, there was not a direct measure of accuracy. Instead, the informativeness of the cues that instantiated heuristics was examined. In most conditions, when Soldiers reported hypotheses early, they were in response to high-value cues. Thus, we can be reasonably confident that Soldiers were generating plausible, well-informed hypotheses. The one exception to this was individual Soldiers generating hypotheses under time pressure. In these cases, Soldiers generated hypotheses based on less information but not necessarily on high-value information. In fact, in scenarios with time pressure and an early high-value cue, nearly one third of hypotheses generated early occurred prior to the appearance of the high-value cue. Encouraging, however, is that multiple factors can potentially mitigate faulty hypothesis generation. First, individual disposition can influence whether Soldiers working individually employ heuristics. Soldiers high in the need for cognitive closure may be less susceptible to engaging in quick, efficient, but risky hypothesis generation. In addition, working in groups may mitigate the tendency to employ heuristics. Soldiers working in groups did not appear as likely as individuals to base their hypotheses on suboptimal information, particularly when under time pressure. To be sure, they were generally less efficient than individual Soldiers, but the influence of group dynamic and group decision processes may have protected Soldiers against potentially biased or faulty hypothesis generation.

This study suffered from two methodological limitations. First, as noted previously, familiarity and cue order were confounded, thus limiting our ability to observe potential interactions between them. This was not devastating, in that the influence of one variable within levels of the other could still be analyzed, albeit with reduced power. In future designs, fully crossing all independent variables would help alleviate this issue. Second, Soldiers were asked to use two decision scales across the familiarity of the decision context. In familiar contexts, Soldiers initially reported on a trinary scale (e.g., low, moderate, or high threat), whereas in unfamiliar contexts, Soldiers initially reported on a binary scale (admit or divert). Although Soldiers reported qualitatively, thus allowing us to code their hypotheses on the same scale, there is a limitation in comparing the values of initial threat ratings, as well as increases in threat ratings, between familiarity conditions. Again, this is not devastating, as our primary focus was on how many images Soldiers would evaluate before abandoning an initial hypothesis, not on the absolute values of initial or subsequent threat ratings. The scenarios were designed to begin at relatively neutral threat levels and change drastically with the addition of high-value cues. Further, because they would remain neutral with the addition of low-value cues, we are reasonably confident that our measures of heuristic usage are valid. Nevertheless, the threat rating scales for future designs would need some revision.

The experiments presented here represent one step toward understanding how decision environments influence the way Soldiers use heuristics to generate hypotheses individually and collectively. The findings imply that as Soldiers perceive their environments and the demands of their tasks differently, they may also assess those environments differently. Different assessments can lead to considering different courses of action that can directly impact Soldier safety and mission success. Therefore, it is critical to better understand the relative influences of hypothesis generation in operational environments. In addition to exploring the factors addressed here, future research should explore the influence of training and experience on the relationships between environmental conditions, decision tasks, and cognitive processes.

References

- Beach, L. R. (1990). *Image theory: Decision making in personal and organizational contexts*. New York: John Wiley & Sons.
- Cohen, M. S., Freeman, J. T., & Wolf, S. (1996). Metacognition in time stressed decision making: Recognizing, critiquing and correcting. *Human Factors*, 38, 206-219.
- Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.
- Fisher, S. D., Gettys, C. F., Manning, C., Mehle, T., & Baca, S. (1983). Consistency checking in hypothesis generation. *Organizational Behavior and Human Performance*, 31, 233-254.
- Garcia-Retamero, R., & Dhimi, M. K. (2009). Take-the-best in expert-novice decision strategies for residential burglary. *Psychonomic Bulletin & Review*, 16(1), 163-169.
- Gettys, C. F., Manning, C., Mehle, T., & Fisher, S. (1980). *Hypothesis generation: A final report of three years of research*. Final report prepared under Contract N00014-77-C-0615 for Office of Naval Research Engineering Psychology Programs). Norman, OK: Department of Psychology, University of Oklahoma.
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Sciences*, 1, 107-143.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451-482.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, 103(4), 650.
- Hammond, K. R. (1988). *Judgment and decision making in dynamic tasks*. Colorado University at Boulder Center for Research on Judgment and Policy.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.) *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 49-81). New York: Cambridge Univ. Press.
- Klein, G. A. (1993). *A recognition-primed decision (RPD) model of rapid decision making*. Ablex Publishing Corporation.
- Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking forward. In C. E. Zsombok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 285-292). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Klein, G. A., & Brezovic, C. P. (1986). Design engineers and the design process: Decision strategies and human factors literature. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 30, No. 8, pp. 771-775). SAGE Publications.

- Latane, B. (1981). The psychology of social impact. *American Psychologist*, 36(4), 343.
- Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37(6), 822.
- Laughlin, R. P., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology*, 22, 177–189.
- Laughlin, P. R., & Shippy, T. A. (1983). Collective induction. *Journal of Personality and Social Psychology*, 45(1), 94.
- Lee, T. W., Mitchell, T. R., Holtom, B. C., McDaniel, L. S., & Hill, J. W. (1999). The unfolding model of voluntary turnover: A replication and extension. *Academy of Management Journal*, 42, 450-462.
- Lee, T. W., Mitchell, T. R., Wise, L., & Fireman, S. (1996). An unfolding model of voluntary employee turnover. *Academy of Management Journal*, 39, 5-36.
- Leins, D. A., Zimmerman, L. A., Marcon, J., Mueller, R., Singer, J. T., & Vowels, C. L. (2013). Assessing threat detection scenarios through hypothesis generation and testing. (ARI Technical Report 1347). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes*, 69, 149-163.
- Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, 222, 250-259. doi: 10.1016/j.jneumeth.2013.10.024.
- Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1166.
- Pachur, T., & Marinello, G. (2013). Expert intuitions: How to model the decision strategies of airport customs officers? *Acta Psychologica*, 144(1), 97-103.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.
- Reimer, T., & Hoffrage, U. (2003). Information aggregation in groups: The approach of simple group heuristics (SIGH). In *Proceedings of the 25th annual conference of the Cognitive Science Society* (pp. 982-987). Erlbaum.
- Reimer, T., & Katsikopolous, K. V. (2004). The use of recognition in group decision making. *Cognitive Science*, 28, 1009-1029.

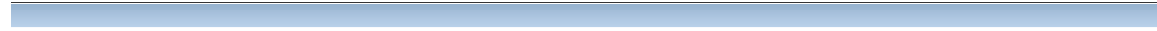
- Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. Todd, & The ABC Research Group (Eds.). *Simple heuristics that make us smart* (pp. 141-167). New York: Oxford University Press.
- Roets, A. & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the need for cognitive closure scale. *Personality and Individual Differences*, 50, 90-94.
- Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement*, 55(5), 818-831.
- Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: an effort-reduction framework. *Psychological Bulletin*, 137, 207-222.
- Shanteau, J. (1992a). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, 53, 252-266.
- Shanteau, J. (1992b). How much information does an expert use? Is it relevant? *Acta Psychologica*, 81, 75-86.
- Simon, H. A. (1957). *Models of man; social and rational*. Oxford, England: Wiley.
- Sorkin, R. D., Luan, S., & Itzkowitz, J. (2004) Group decision and deliberation: A distributed detection process. In D. J. Koehler and N. Harvey (Eds.) *Blackwell Handbook of Judgment and Decision Making* (pp. 464-484). Malden, MA: Blackwell Publishing Ltd.
- Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, 115, 155-185.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, 67(6), 1049.

Appendix A

Scenario Examples

The following images are screen captures of stimuli presented in the PEBL application.

Image 1. General Instructions



You are about to view several images of situations that you will assess. You will receive specific instructions with each image. These instructions will indicate the type of assessment you need to make and how quickly you must make it. You will have 12 seconds to read each instruction.

You will see an image and then write your initial assessment of that image, including the reasons for your assessment, and then rate your confidence in your assessment on a scale of 0-100. Then the image will change over time. Information will be added to the image that may or may not help you further assess the situation.

You will press and hold the LEFT MOUSE BUTTON if and when your assessment changes. You will then type a brief description of your assessment, and your confidence. You will press ENTER after typing your response and receive instructions for assessing the next situation.

You will now practice interacting with the test program. LEFT MOUSE CLICK to continue.

Image 2. Familiar context: Initial image & decision task



Your squad is tasked with classifying a route through a valley. Your squad leader has asked you to report on the threat level of classifying this route. You must report back on how safe it is to continue through this valley.

Image 3. Familiar context: Initial hypothesis entry screen



How safe it is to continue down this road?

A large, empty gray rectangular box intended for the user to type their assessment of the situation.

Please briefly describe your assessment of this situation. Type your answer, then press ENTER

Image 4. Familiar context: High-value cue presented in serial position two



IF your assessment changes, press and hold the LEFT MOUSE BUTTON

Note. The initial image depicted the valley minus the portions of the dirt road, the vehicle, and the red arrow (see Image 3 on page A-3).

Image 5. Familiar context: Second hypothesis entry screen

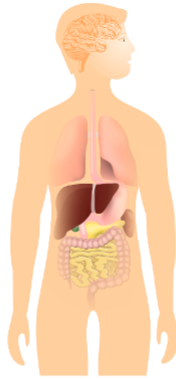


How safe it is to continue down this road?

A large, empty gray rectangular box intended for the user to type their assessment of the situation.

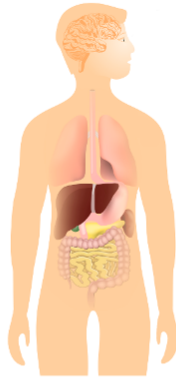
Please briefly describe your assessment of this situation. Type your answer, then press ENTER

Image 6. Unfamiliar context: Initial image & decision task



This patient is 62 years old, divorced, and reports weight loss, dropping from 190 to 185 lbs, but cannot attribute the weight loss to anything specific. He recently had a dentist visit for a minor procedure. He currently has a mild fever.

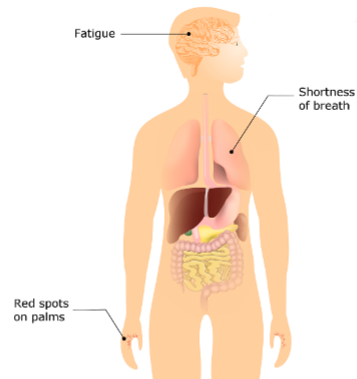
Image 7. Unfamiliar context: Initial hypothesis entry screen



Admit or Divert this patient?

Please briefly describe your assessment of this situation. Type your answer, then press ENTER

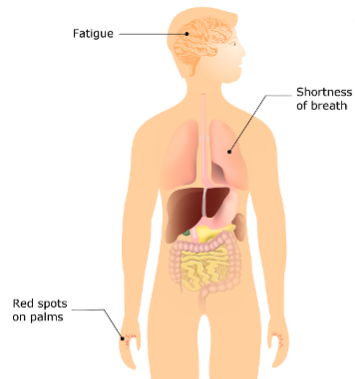
Image 8. Unfamiliar context: High-value cue presented in serial position three



IF your assessment changes, press and hold the LEFT MOUSE BUTTON

Note. The high-value cue in this image is the collection of red spots on the patient's palms.

Image 9. Unfamiliar context: Second hypothesis entry screen



Admit or Divert this patient?

Please briefly describe your assessment of this situation. Type your answer, then press ENTER

Appendix B
Questionnaires and Scales

Demographic Questionnaire

1. Please enter your time in service, in years.
2. Please enter your current rank.
3. Please enter your time in current rank, in months.
4. Please enter your current MOS.
5. Please enter your age.
6. Have you ever deployed? Please answer with 'yes' or 'no.'
 - a. If yes, how many times have you deployed?
7. Please enter the location of your most recent deployment (city or cities, and country).
8. Please enter your MOS at the time of your most recent deployment.
9. How often did you go 'outside the wire' on your most recent deployment? Please answer with: (0- never, 1- less than once a month, 2- once a month, 3- more than once a month, 4- Once a week, 5- More than once a week, 6- Every day).
10. Please describe any training you have received that improved your ability to detect threats and indicate the approximate date of the training month/year.
11. Please enter how long you have been with your current unit/squad. Enter your response in days, months, or years.
12. Have you been deployed with this current unit/squad? Yes/No

Decision-Making Style Scale

Scored on a 1-5 scale (1 = Strongly Disagree, 2 = Somewhat Disagree, 3 = Neither Agree nor Disagree, 4 = Somewhat Agree, 5 = Strongly Agree)

“Listed below are statements describing how individuals go about making important decisions. Please indicate how much you agree with each statement.”

1. I double-check my information sources to be sure I have the right facts before making decisions.
2. I make decisions in a logical and systematic way.
3. My decision making requires careful thought.
4. When making a decision, I consider various options in terms of a specific goal.
5. When I make decisions, I rely upon my instincts.
6. When making decisions, I tend to rely on my intuition.
7. I generally make decisions that feel right to me.
8. When I make a decision, it is more important for me to feel the decision is right than to have a rational reason for it.
9. When I make a decision, I trust my inner feelings and reactions.
10. I often need the assistance of other people when making important decisions.
11. I rarely make important decisions without consulting other people.
12. If I have the support of others, it is easier for me to make important decisions.
13. I use the advice of other people in making my important decisions.
14. I like to have someone to steer me in the right direction when I am faced with important decisions.
15. I avoid making important decisions until the pressure is on.
16. I postpone decision making whenever possible.
17. I often procrastinate when it comes to making important decisions.
18. I generally make important decisions at the last minute.
19. I put off making many decisions because thinking about them makes me uneasy.
20. I generally make snap decisions.
21. I often make decisions on the spur of the moment.
22. I make quick decisions.
23. I often make impulsive decisions.
24. When making decisions, I do what seems natural at the moment.

Need for Cognitive Closure Scale

Scored on a 1-6 scale (1 = Strongly Disagree, 2 = Moderately Disagree, 3 = Slightly Disagree, 4 = Slightly Agree, 5 = Moderately Agree, 6 = Strongly Agree)

“Read each of the following statements and decide how much you agree with each according to your beliefs and experiences.”

1. I don't like situations that are uncertain.
2. I dislike questions which could be answered in many different ways.
3. I find that a well ordered life with regular hours suits my temperament.
4. I feel uncomfortable when I don't understand the reason why an even occurred in my life.
5. I feel irritated when one person disagrees with what everyone else in a group believes.
6. I don't like to be with people who are capable of unexpected actions.
7. I don't like to into a situation without knowing what I can expect from it.
8. I dislike it when a person's statement could mean many different things.
9. I find that establishing a consistent routine enables me to enjoy my life more.
10. I enjoy having a clear and structured mode of life.
11. I do not usually consult many different options before forming my own view.
12. I dislike unpredictable situations.
13. When I have made a decision, I feel relieved.
14. When I am confronted with a problem, I'm dying to reach a new solution very quickly.
15. I would quickly become impatient and irritated if I would not find a solution to a problem immediately.

Modified Need for Cognitive Closure Scale from Roets and Van Hiel (2011).

Original Need for Cognitive Closure Scale developed and validated by Webster and Kruglanski (1994).

Appendix C

Hypothesis Scoring Examples

Hypotheses scored as 0 (low or minimal threat/urgency):

- Threat detection (familiar) scenario: “No immediate threat. Threat level low.”
- Medical diagnosis (familiar) scenario: “Divert. Headaches can be many minor things.”

Hypotheses scored as 1 (moderate threat/urgency):

- Threat detection scenario: “Medium threat level. Potential hiding spots along road for enemy scouts.”
- Medical diagnosis scenario: “Divert. I would schedule a follow up appointment in a couple of days.”

Hypotheses scored as 2 (high threat/urgency):

- Threat detection scenario: “The threat level is high because there is multiple hiding places along the road. Perfectly suited for an ambush.”
- Medical diagnosis scenario: “Admit. Could be about to have a heart attack.”